Project Report:

An analysis of the union wages data: GLM's, GAM's and JAGS

STAT 260: Nonparametric Statistics

Henrique Ap. Laureano
ID 158811

`/KAUST/CEMSE/STAT`

Spring Semester
2018

Contents

Data and goals

In this project we study the `union wages data` (`trade.union` data in the `R` [1] package `SemiPar` [2]). In this dataset we have the record of 534 U.S. workers. The main variable, the target, is a qualitative variable saying if the worker is a trade union membership (or not).

Together with this we have more ten variables/features, briefly described in Table 1. From this ten features, four are quantitative and 6 are qualitative. To know the behaviour of this features some descriptive analysis is performed, see Figure 1 and Figure 2.

Table 1: Features description of the `union wages data`.

Feature	Description
`union.member`	trade union membership indicator (target variable): 0 = non-member, 1 = member
`years.educ`	number of years of education
`south`	indicator of living in southern region of U.S.A.
`female`	gender indicator: 0 = male, 1 = female
`years.experience`	number of years of work experience
`wage`	wages in dollars per hour
`age`	age in years
`race`	1 = black, 2 = Hispanic, 3 = white
`occupation`	1 = management, 2 = sales, 3 = clerical, 4 = service, 5 = professional, 6 = other
`sector`	0 = other, 1 = manufacturing, 2 = construction
`married`	indicator of being married: 0 = unmarried, 1 = married

In Figure 1 we see the scatterplots and correlations, two-by-two, for the four quantitative features in the `union wages data`. From the six scatterplots generated we just see a clear behavior - relation in this case - in two of them, `years.educ` vs. `wage` and `years.experience` vs. `age`. In the first we see a, clear, linear increasing behaviour - that makes all the sense, since with more years of education we expect a bigger salary. However, we also see a high variability in this salaries, which is reflected in the positive, but not high, linear correlation of 0.38. Already in the `years.experience` vs. `age` scatterplot we see a very strong linear relation, reflected in the correlation of 0.98 (almost perfect), which shows a small variability. Again the observed behaviour makes all the practical and expected sense. In the others scatterplots the aleatority and variability also makes sense - in general is hard to define a common and exactly behaviour between features.

Now, in Figure 2 we have the frequencies for each level of the categorical features in the `union wages data`. First, we see that the target variable is unbalanced, with more than 80% of the observations corresponding to non-members of the trade union. By this Figure we see that most of the workers, in the dataset, live in the southern region of U.S. (`south` feature), are males (`female` feature), white (`race` feature), have a job that does not fit into the other

(five) available categories - management, sales, clerical, service and professional (`occupation` feature), works on the construction `sector` and are `married`. This is the average profile of the 534 U.S. workers in the `union wages data`.

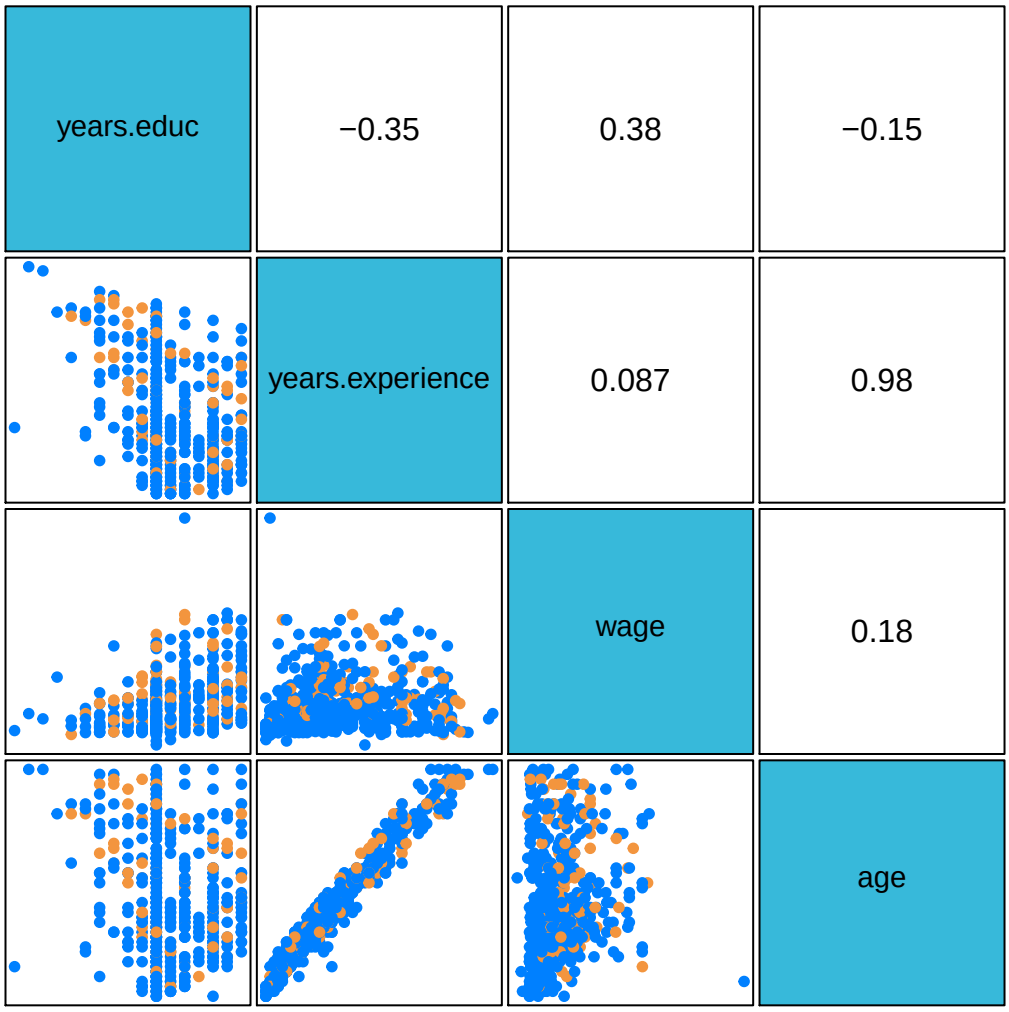

Figure 1: Scatterplot lower triangular matrix with colors representing the trade union membership status and, correlation upper triangular matrix for the quantitative features in the `union wages data`.

Using this features, described in Table 1, the goals here are

- Test/apply, some models to see how good they are to classify the target feature, i.e., classify, given the ten features, if the worker is a trade union membership or not.
- See which features are important, statistically significant, to discriminate the worker in a member or not of the trade union.
- Study the behaviour of the quantitative features. They have a linear behaviour? If they don't have and we fit a model considering a linear relationship, we will get bad results?

The models and algorithms used for this task are described in the next section, together with some extra informations about the analysis procedure.

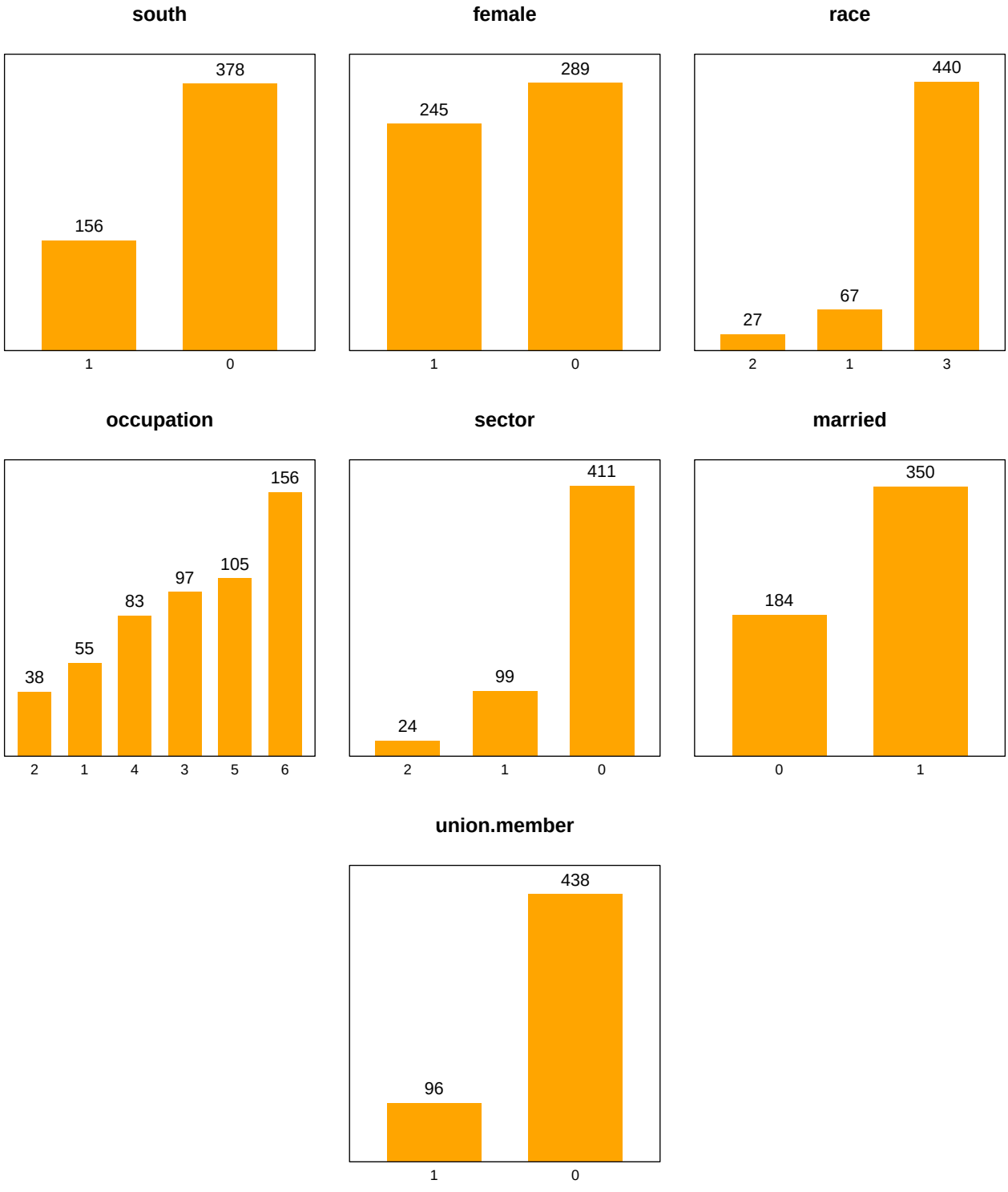

Figure 2: Barplots for the qualitative features in the `union wages data`, with original labels.

Methods

To classify, given the ten features, if the worker is a trade union membership or not, five, six model are used. They are

- Four generalized linear models with a Bernoulli response and with different link functions (logit, probit, cauchit and complementary log-log);
- Four generalized additive models with integrated smoothness estimation for the quantitative features (to reach, approach, the last point mentioned in the end of the previous section) and, with a Bernoulli response and different link functions (logit, probit, cauchit and complementary log-log);
- A bayesian model fitted via JAGS (Just Another Gibbs Sampler) - `http://mcmc-jags.sourceforge.net/`. To be clear in the explanation, the justification for this last approach will be given in the results section - where the approach introduction will be more didactical and logical.

All the analysis are performed using the `R` [1] language and environment for statistical computing. To take advantage of the most efficient available algorithm versions we use some `R` libraries where the algorithms are implemented. A brief description of the algorithms is given below, always mentioning the corresponding `R` library where the algorithm is implemented.

To do feature selection and test the significance of the features we use the Akaike Information Criterion (AIC). Given a collection of models, the AIC estimates the quality of each model, relative to each of the other models. Thus, AIC provides a means for model selection. The AIC value of a given model is the following:

$$\text{AIC} = 2p - 2\log\widehat{L}.$$

With $\widehat{L}$ being the maximum value of the likelihood function for the model and p being the number of estimated parameters in the model.

For more details, a simple and intuive material can be found in this Wikipedia page `https://en.wikipedia.org/wiki/Akaike_information_criterion`.

Generalized Linear regression Model (GLM) with Bernoulli response

The GLM is a flexible generalization of ordinary linear regression that allows responses with error distribution models different from the normal distribution. The GLM generalizes linear regression by allowing the linear model to be related to the response via a link function. In a GLM each outcome Y of the response is assumed to be generated from a particular distribution in the exponential family, a large range of probability distributions. The mean, μ, of the distribution depends on the features, X, through:

$$\text{E}(Y) = \boldsymbol{\mu} = g^{-1}(X\boldsymbol{\beta}),$$

where $\text{E}(Y)$ is the expected value of Y; $X\boldsymbol{\beta}$ is the linear predictor, a linear combination of unknown parameters $\boldsymbol{\beta}$; g is the link function. The unknown parameters, $\boldsymbol{\beta}$, are typically estimated with maximum likelihood.

When the response data, Y, are binary (taking on only values 0 and 1), the distribution function is generally chosen to be the Bernoulli distribution and the interpretation of μ_i is then

the probability, p, of Y_i taking on the value one. The logit is the canonical link function and when used the resulting model is called of logistic regression. However, other link function can be used. The four most popular link functions, and used here, are:

- Logit function: $g(p) = \ln\left(\frac{p}{1-p}\right)$;
- Probit or inverse Normal function: $g(p) = \Phi^{-1}(p)$;
- Cauchit function: $g(p) = \tan\left(\pi p - \frac{\pi}{2}\right)$;
- Complementary log-log function: $g(p) = \log(-\log(1-p))$.

More details about GLM can be see, for example, in `https://en.wikipedia.org/wiki/Generalized_linear_model` and, of course, in the main reference of the subject, [3].

Generalized Additive Model (GAM) with Bernoulli response

The GAM is a generalized linear model in which the linear predictor depends linearly on unknown smooth functions of some predictor features, and interest focuses on inference about these smooth functions. The model relates a univariate response, Y, to some predictors, $\mathbf{x}_i$. An exponential family distribution is specified for Y along with a link function g relating the expected value of Y to the predictors via a structure such as

$$g(\mathrm{E}(Y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \cdots + f_m(x_m).$$

The functions f_i may be functions with a specified parametric form (for example a polynomial, or an un-penalized regression spline of a feature) or may be specified non-parametrically, or semi-parametrically, simply as 'smooth functions'. A typical GAM might use a scatterplot smoothing function, such as a locally weighted mean, for $f_1(\mathbf{x}_1)$, and then use a factor model for $f_2(\mathbf{x}_2)$. This flexibility to allow non-parametric fits with relaxed assumptions on the actual relationship between response and predictor, provides the potential for better fits to data than purely parametric models, but arguably with some loss of interpretability.

The original GAM fitting method estimated the smooth components of the model using non-parametric smoothers (for example smoothing splines or local linear regression smoothers) via the backfitting algorithm. Backfitting works by iterative smoothing of partial residuals and provides a very general modular estimation method capable of using a wide variety of smoothing methods to estimate the $f_j(x_j)$ terms. A disadvantage is that it is difficult to integrate with the estimation of the degree of smoothness of the model terms, so that in practice the user must set these, or select between a modest set of pre-defined smoothing levels.

If $f_j(x_j)$ are represented using smoothing splines then the degree of smoothness can be estimated as part of model fitting using generalized cross validation, or by REML. Many modern implementations are built around the reduced rank smoothing approach, because it allows well founded estimation of the smoothness of the component smooths at comparatively modest computational cost, and also facilitates implementation of a number of model extensions. At its simplest the idea is to replace the unknown smooth functions in the model with basis expansions.

When the response Y is binary the approach idea is the same that with GLM's, and the available, and used, link functions are the same four mentioned above in the GLM subsection.

In R the recommended package for GAM's, and used here, is the mgcv (mixed gam computational vehicle) [4] which is based on the reduced rank approach with automatic smoothing parameter selection.

More about GAM's can be found in this, very nice, Wikipedia page https://en.wikipedia.org/wiki/Generalized_additive_model and, mainly, in this two references, [4] and [5].

Just Another Gibbs Sampler (JAGS)

JAGS is Just Another Gibbs Sampler. It is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation not wholly unlike BUGS (https://www.mrc-bsu.cam.ac.uk/software/bugs/). JAGS was written with three aims in mind:

- To have a cross-platform engine for the BUGS language;
- To be extensible, allowing users to write their own functions, distributions and samplers;
- To be a platform for experimentation with ideas in Bayesian modelling.

The main advantage of JAGS in comparison to the members of the original BUGS family (WinBUGS and OpenBUGS) is its platform independence. It is written in C++, while the BUGS family is written in Component Pascal, a less widely known programming language. In addition, JAGS is already part of many repositories of Linux distributions. JAGS can be used via the command line or run in batch mode through script files. This means that there is no need to redo the settings with every run and that the program can be called and controlled from within another program (e.g. from R via rjags [6], as we did here).

The main references about JAGS can be found in https://sourceforge.net/projects/mcmc-jags/ and in the JAGS user manual https://martynplummer.wordpress.com/2017/06/28/new-manual/.

Results

The fitted GLM's, with ten features each, is represented by the linear predictor in Equation 1.

$$\begin{aligned} g(p_i) &= \beta_0 + \beta_1 \texttt{years.educ}_i + ... + \beta_{10} \texttt{married}_i, \\ \texttt{union.member}_i &\sim \text{Bernoulli}(p_i), \quad i = 1, \dots 534. \end{aligned} \tag{1}$$

where g represents the link function (logit, probit, cauchit or complementaty log-log) and p_i is the probability of trade union membership. The fitted GAM's, also with ten features each, is represented by the linear predictor in Equation 2.

$$\begin{aligned} g(p_i) &= \beta_0 + f_1(\texttt{years.educ}_i) + ... + f_4(\texttt{age}_i) + \beta_1 \texttt{female}_i + ... + \beta_6 \texttt{married}_i, \\ \texttt{union.member}_i &\sim \text{Bernoulli}(p_i), \quad i = 1, \dots 534. \end{aligned} \tag{2}$$

where we have four quantitative features, thus we have four smooth functions.

Doing features selection in this GLM's and GAM's via AIC we arrive in the models presented in Table 2, where we see the keeped (with a checkmark) and dropped (without a checkmark) features, and the AIC value of each final model.

Table 2: Remaining features in each model (specified by the link function and used approach, GLM or GAM) after features selection by AIC, and the obtained AIC in each final model (smallest AIC in bold).

Feature	Logit		Probit		Cauchit		Comp. log-log	
	GLM	GAM	GLM	GAM	GLM	GAM	GLM	GAM
years.educ								
south	✓	✓	✓	✓	✓		✓	
female	✓	✓	✓	✓	✓		✓	✓
years.experience								
wage	✓	✓	✓	✓	✓	✓	✓	✓
age	✓	✓	✓	✓	✓	✓	✓	✓
race	✓	✓	✓	✓	✓	✓	✓	✓
occupation	✓	✓	✓	✓	✓	✓	✓	✓
sector								
married	✓		✓				✓	
AIC	452.87	**436.67**	453.34	437.83	454.91	**436.55**	452.77	436.71

A nice and interesting thing to notice in Table 2 is the fact that we have two AIC's virtually equals, but in one model (GAM with cauchit link function) we have four features, and in the other model (GAM with logit link function) we have six features. To see what's happening we have Table 3 and Table 4 with the summaries of this two models, respectively.

The main thing to mention about this summaries is that with the cauchit link function the standard erros are hugh. Before take any further conclusion we look to the residuals and goodness-of-fit of this models. We present this in Figure 3. There we see the dispersion of the Pearson and Deviance residuals, and the ROC curves for each model. The Receiver Operating Characteristic curve, i.e. the ROC curve, illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the specificity, true negative rate, against the sensitivity, true positive rate, at various threshold/cutoof settings.

More details about can be see, for example, in `https://en.wikipedia.org/wiki/Receiver_operating_characteristic`. In `R` the main implementation of the ROC curve is found in the `pROC` library [7].

In Figure 3 we see that both models don't present very good (values centraded in zero and around -3 and 3) residuals and very high Area Under the Curve (AUC), that is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. However, the results are also not so bad. The AUC's are bigger than 0.75 in both models and the sensitivities and specificities are bigger than 0.65. In the residual dispersions, mainly with the model with logit link function, we see that most part of the points are centrated in zero and around -3 and 3, with very few exceptions.

Table 3: Summary of the GAM with cauchit link function, the best model - by the AIC.

Parametric coefficients	**Estimate**	**Std. Error**	t-**value**	p-**value**
Intercept[1]	-7.0130	3.1102	-2.2548	0.0241
`age`	0.0355	0.0144	2.4743	0.0133
`race` (Hispanic)	-0.3775	0.8884	-0.4249	0.6709
`race` (white)	-0.9580	0.4352	-2.2014	0.0277
`occupation` (sales)	-2.3917	7.9952	-0.2991	0.7648
`occupation` (clerical)	2.3330	3.2347	0.7212	0.4708
`occupation` (service)	5.2263	3.0506	1.7132	0.0867
`occupation` (professional)	4.3119	3.0389	1.4189	0.1559
`occupation` (other)	5.2294	3.0328	1.7243	0.0847
Smooth terms	**edf**	**Ref.df**	F-**value**	p-**value**
s(`wage`)	3.4255	4.4079	22.1916	0.0003

[1] Intercept is the reference level, equivalent to a worker black (`race`) and working (`occupation`) in management.

Table 4: Summary of the GAM with logit link function, the 2nd best model - by the AIC.

Parametric coefficients	**Estimate**	**Std. Error**	t-**value**	p-**value**
Intercept[1]	-3.6292	0.8164	-4.4454	< 0.0001
`south` (yes)	-0.4736	0.3047	-1.5540	0.1202
`female` (yes)	-0.5001	0.3018	-1.6572	0.0975
`age`	0.0275	0.0110	2.4885	0.0128
`race` (Hispanic)	0.0228	0.6403	0.0356	0.9716
`race` (white)	-0.7528	0.3530	-2.1327	0.0329
`occupation` (sales)	-0.3412	1.2014	-0.2840	0.7764
`occupation` (clerical)	1.1197	0.7409	1.5113	0.1307
`occupation` (service)	2.3392	0.7005	3.3395	0.0008
`occupation` (professional)	1.8141	0.6567	2.7623	0.0057
`occupation` (other)	2.4219	0.6520	3.7145	0.0002
Smooth terms	**edf**	**Ref.df**	F-**value**	p-**value**
s(`wage`)	2.7437	3.5144	25.1623	< 0.0001

[1] Intercept is the reference level, equivalent to a worker not from the `south`, male, black and working (`occupation`) in management.

Given the behaviours presented in Figure 3, and together with the fact that is very hard to get very nice residual results with binary data, we are able to say that we get a better result with the GAM model using the logit link function, and that the residual analysis and goodness-of-fit of this model are satisfatory. Therefore, from the eight models presented in Table 2 the best model is the GAM with logit link function.

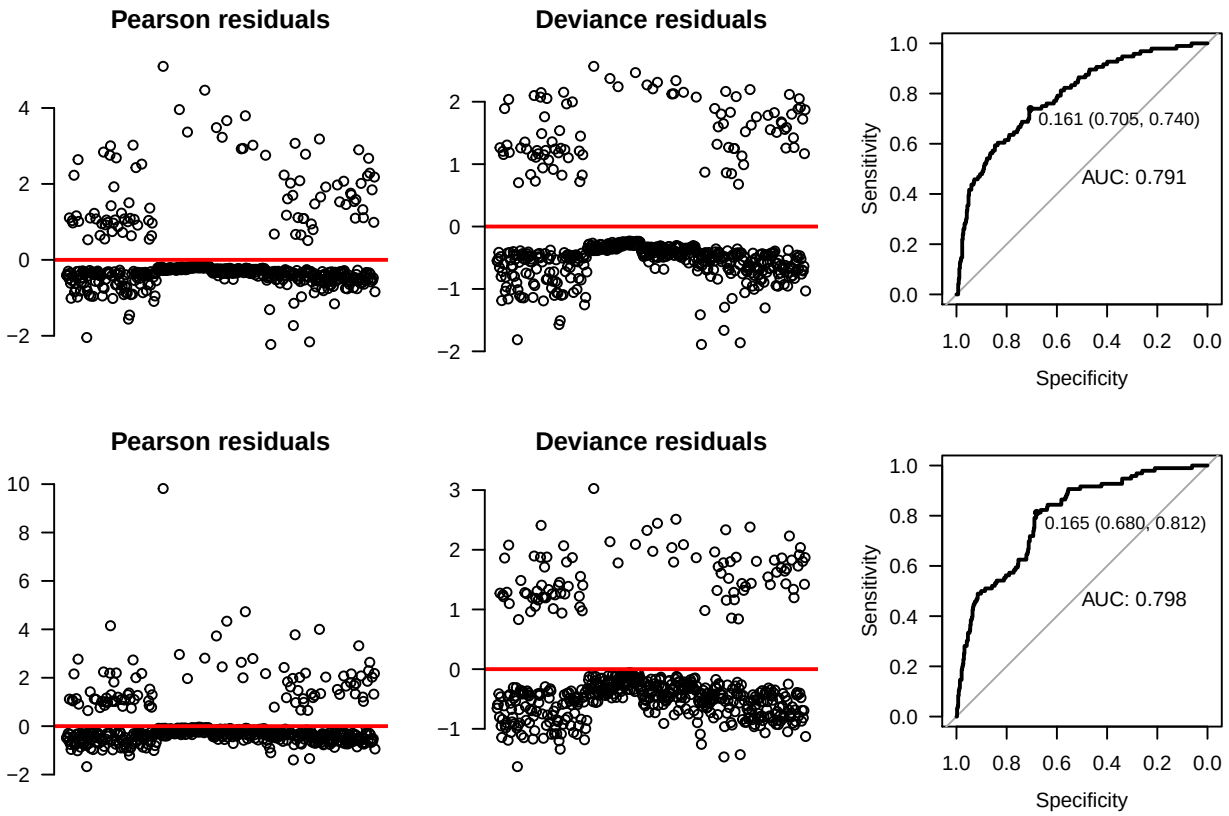

Figure 3: Residual analysis and goodness-of-fit. First line: GAM with cauchit link function. Second line: GAM with logit link function. Graphs: Dispersion of the Pearson and Deviance residuals, in the left and in the center, respectively. ROC curve in the right, with AUC value, cutoff, specificity and sensitivity.

In the GAM model with logit link function we have six features, as we saw in Table 2, but from this six just two are quantitative, `wage` and `age`. As we can see in Table 4, in this model we have only one smooth term, the feature `wage`. In all GAM's we started considering all the four quantitative features as smooth, neverthless, by variable selection via AIC we saw that in all models (GAM's and GLM's) the feature `wage` is the only one that present a non-linear relation. All others present a linear relation and, from this other three only the feature `age` is statiscally significant. The estimated smooth function for the feature `wage` by the model with logit link function is presented in the Figure 4 together with a confidence band of two standard errors.

We see in this Figure that we have only one worker with a `wage` bigger than 30, and because of this in that part the standard error is hugh - the uncertainty is big. Before this point the uncertainty is very small and we see that the smooth function increase until a `wage` close to 15, from this point forward the function is almost constant with the uncertainty increasing conform as we are having less observations.

A question of curiosity and to see what happens/the behavior, we can fit a Bayesian model, simulate from the model and take a look on some samples from the posterior.

As the interest at this point is about this feature with a non-linear behaviour described in Figure 4, we fit a Bayesian model using JAGS (`http://mcmc-jags.sourceforge.net/`) and only considering the feature `wage` - with a smooth function. The model is represented by the

linear predictor presented Equation 3.

$$\text{logit}(p_i) = f(\texttt{wage}_i), \quad \texttt{union.member}_i \sim \text{Bernoulli}(p_i), \quad i = 1, \ldots 534. \tag{3}$$

where p_i is the probability of trade union membership.

In this model we consider the logit link function, since is the link function used in the best (Table 4) of the eight fitted models (GLM's and GAM's). The prior considered for the smoothing parameter is the default prior available in JAGS module. In this case, a flat Gamma.

After the model fitting we are able to simulate from the model. In Figure 5 we present the modelled probability of union membership against `wages`, with a credible interval. A sample of twenty curves from the posterior is added to provide a better understanding of the smooth shape range. We can see in this Figure that even considering in the prediction a interval between 0 and 30 the interval is wide at high `wages`. Now we can also see better the behavior, how - in which shape - the function increase until a `wage` of 15, and how it decrease after this value. With the twenty smooth curves draw from the posterior we are able to examine better the variability in the smooth shape. With very few exceptions in vey specific points we see the samples going outside the credibility interval.

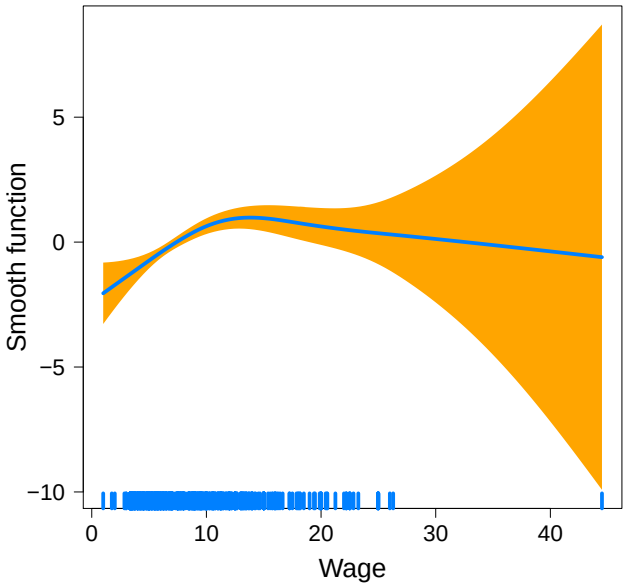

Figure 4: Smooth function for the feature `wage` with confidence band of two standard errors, obtained with the GAM model with logit link function.

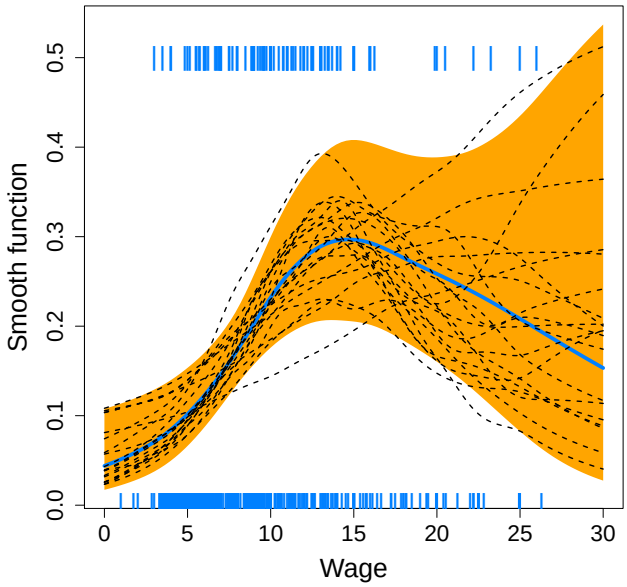

Figure 5: Probability of union membership against `wages`, with a credible interval and twenty curves from the posterior.

Now, with all this results in mind we are able to reach the conclusions in the next section.

Conclusions

In Table 2 we saw an agreement between the models for most of the features. The features `wage`, `age`, `race` and `occupation` are present in all of them. The features `years.educ`,

years.experience and sector are dropped from all models.

First, let's talk about the quantitative features. years.educ and years.experience are not statistically significants to discriminate the workers in members of the trade union. In Figure 1 we saw a very high correlation between the features years.experience and age, and during the features selection process we confirm that they have the same discriminative power, and that in the presence of one, the other is extremely not significant. Fitting different models one time with one, and other time with the other, we reach the conclusion that the age feature is slightly more significant.

All this four features in the GAM's started as smoothing functions, however, as the features selection showed, only wage have a non-linear behaviour. Even considering as linear, wage is significant in the GLM's. However, as the AIC's show in Table 2, is better/more adequate to consider this feature as smooth.

Talking about the qualitative features now. In the final/best obtained model - GAM with logit link function, we don't have the features sector and married, i.e., to discriminate the workers in members of the trade union, the information about work sector and civil status are not important, statistically significant. We measured the effect of this features and the result was presented in Table 4. Workers that don't live in the south; males; Hispanics; and with an occupation that is not in management, sales, clerical, service and professional field, have the biggest probability of trade union membership.

The different used approachs returned coherent results and the last of this approachs, the Bayesian model via JAGS, presented very interesting results, as we saw in Figure 5. In the end, the approachs complemented each other and enabled us to tell a 'story' in a coherent and interesting flow. A last detail. In the smooth functions we started considering basis of dimension 10 - the default option of the function. After this we tried different values to see with which one we obtain a better result. For the feature wage the best result was obtained with a basis dimension of 20.

References

[1] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. https://www.R-project.org/.

[2] M. Wand, *SemiPar: Semiparametic Regression*, R package version 1.0-4.1 2014. https://CRAN.R-project.org/package=SemiPar.

[3] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, vol. 37 of *Monographs on Statistical and Applied Probability*. Chapman and Hall, second ed., 1989.

[4] S. N. Wood, *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, second ed., 2017.

[5] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. Chapman and Hall/CRC, second ed., 1990. ISBN 978-0-412-34390-2.

[6] M. Plummer, *rjags: Bayesian Graphical Models using MCMC*, R package version 4-6 2016. https://CRAN.R-project.org/package=rjags.

[7] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "proc: an open-source package for r and s+ to analyze and compare roc curves," *BMC Bioinformatics*, vol. 12, p. 77, 2011.

Appendices: R generic code

Reading data

```
library(SemiPar) # ========================= loading dataset library #
data(trade.union, package = "SemiPar") # ========== loading dataset #

# ================ converting qualitative features in R factor class #
trade.union[ , c(2:3, 5, 8:11)] <-
  lapply(trade.union[ , c(2:3, 5, 8:11)], factor)
```

Fitting GLM's

```
library(MASS) # ========= loading library for the stepAIC() function #

# ====== link_function: logit (default), probit, cauchit or cloglog #
glm.linkfun <- glm(union.member ~ .,
                   family = binomial(link = "link_function"),
                   trade.union)

# ================================== performing fatures selection #
glm.linkfun <- stepAIC(glm.linkfun)
```

Fitting GAM's

```
library(mgcv) # == loading mixed gam computational vehicle library #

# ============================================== general model formula #
## ============================== default value for 'basis': 10 ##
formula = union.member ~
  s(years.educ, k = "basis") + south + female +
  s(years.experience, k = "basis") + s(wage, k = "basis") +
  s(age, k = "basis") + race + occupation + sector + married
# ====== link_function: logit (default), probit, cauchit or cloglog #
gam.linkfun <- gam(formula,
                   family = binomial(link = "link_function"),
                   trade.union)

# performing features selection comparing two nested models and... #
anova(gam.linkfun_bigger, gam.linkfun_smaller, test = "Chisq")
```

```
16
17 gam.linkfun <- gam.linkfun_bigger # ======= or gam.linkfun_smaller #
```

Performing residual analysis and goodness-of-fit

```
1  # ================== computing and plotting the person residuals #
2  pearson.model_linkfun <-
3    residuals(model.linkfun, type = "pearson")
4  plot(pearson.model_linkfun)
5
6  # ================== computing and plotting the deviance residuals #
7  deviance.model_linkfun <-
8    residuals(model.linkfun, type = "deviance")
9  plot(deviance.model_linkfun)
10
11 library(pROC) # ============== loading library the roc() function #
12 # ========================= computing and plotting the roc curve #
13 plot.roc(roc(trade.union$union.member, fitted(model.linkfun)))
```

Fitting JAGS

```
1  library(rjags) # ======= loading library that connects JAGS with R #
2  data(trade.union, package = "SemiPar") # ================== data #
3
4  # ========================================== setting up the model #
5  ## ====================== jagam(): function from the mgcv library ##
6  jags.gam <- jagam(union.member ~ s(wage, k = 20),
7                    trade.union,
8                    family = binomial(link = "logit"),
9                    file = "file_name.jags")
10 ##  to look to the model go to the JAGS model specification file: ##
11 ## ========================================== file_name.jags ##
12
13 # ======================== compiling and simulating from the model #
14 union.jags <- jags.model("file_name.jags",
15                          data = jags.gam$jags.data,
16                          inits = jags.gam$jags.ini,
17                          n.chains = 3) # ======= value that I used #
18
19 ## ======================== extracting random samples from the ##
20 ## ================== posterior distribution of the parameters ##
21 samp <- jags.samples(union.jags, c("b", "rho", "mu"),
22                      n.iter = 1e4, # ========== value that I used #
23                      thin = 10) # ============  value that I used #
24 jam <- sim2jam(samp, jags.gam$pregam)
```