Classification Using Bayesian Logistic Regression:

Diabetes in Pima Indian Women

Presented by

Ke Wang

Prasenjit Kapat

Outline

- Objectives and Data description
- Model Setup and choice of priors
- Analysis with Gaussian priors
- Analysis with Laplacian priors
- Conclusions

Objectives

- Classification problem from a bayesian perspective.
- Do not intend to focus on model selection or interpretation.
- Effects on the prediction error rate for different choices of prior for the parameters in the logistic regression model.

Data Description

- Data : Pima Indian Women Diabetes data
- A population of women who were at least 21 years old of Pima Indian heritage and living near Phoenix were chosen. They were tested for diabetes.
- The variables in our model were:

npreg	:	number of pregnencies
glu	:	plasma glucose concentration
bp	:	diastolic blood pressure (mm Hg)
skin	:	triceps skin fold thickness (mm)
bmi	:	body mass index kg/m^2
ped	:	diabetes pedigree function
age	:	Age (years)
response	:	*Yes* or *No* for diabetic according to WHO criteria.

- Total 532 instances of which 200 are in training set.

Basic Model

$$y_i|p_i \sim \text{Bernoulli}(p_i)$$

and assume

$$\text{logit}(p_i) = \log\left[\frac{p_i}{1-p_i}\right] = \boldsymbol{x}_i^T\boldsymbol{\beta}$$

Choice of Priors

- Zellner's g-prior: $p(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, g \times (\boldsymbol{X}^T\boldsymbol{X})^{-1})$
- Jeffrey's Prior : $p(\boldsymbol{\beta}) \propto [\text{det}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{V}(\boldsymbol{\beta})\Delta^2(\boldsymbol{\beta})\boldsymbol{X})]^{1/2}$
- Gaussian prior: $p(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta_0}, \boldsymbol{\Sigma_0})$
- Laplacian prior: $p(\beta_j|\lambda_j) = \frac{\lambda_j}{2}e^{-\lambda_j|\beta_j|}$

Analysis with Gaussian Prior

- $p(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$
- We use empirical Bayesian approach to estimate $\boldsymbol{\beta}_0$ and Σ_0 from the data;
- Based on the derived prior distribution, we obtain the posterior distribution of the regression parameters by MCMC approach;
- Based on the posterior distribution, prediction of the binary response will be made using two prediction rules, and relevant measures of prediction error will be calculated.

Analysis with Gaussian Prior: Continued

- Pridict the presence of diabetes based on the posterior distribution:
  i) For each $q_i^{(j)}$, we draw a $y_i^{(j)} \sim \text{Bernoulli}(q_i^{(j)})$, and let $y_i^{pre} = 1$ if the number of $y_i^{(j)}$'s equal to 1 is bigger than or equal to the number of $y_i^{(j)}$'s equal to 0, i.e. the proportion of $y_i^{(j)}$'s equal to 1 exceeds 1/2, otherwise let $y_i^{pre} = 0$. We will refer to this prediction rule as rule i);
  ii) Let $y_i^{pre} = 1$ if the median of the $q_i^{(j)}$'s is bigger than or equal to 1/2, otherwise set $y_i^{pre} = 0$. We will refer to this prediction rule as rule ii).

Analysis with Laplacian Prior

- Optimization based approach: minimise $-l(\boldsymbol{\beta})$ to obtain the *MAP* estimator.
- Initial prior: $[\beta_j|\tau_j] \sim N(0, \tau_j)$
- prior on hyperparameters: $[\tau_j|\gamma_j] \sim Exp(\frac{1}{2}\gamma_j)$
- integrating out τ_j, we have $[\beta_j|\lambda_j] \sim DE(0, \lambda_j)$
- posterior negative-log-density of $\boldsymbol{\beta}$ is given by,

$$-l(\boldsymbol{\beta}) = \left[\sum_{i=1}^{n} \log\left(1 + exp\{-(\beta_0 + \boldsymbol{x}'\boldsymbol{\beta})y_i\}\right)\right]$$

$$+ \sum_{j=0}^{d-1} (\log 2 - \log \lambda_j + \lambda_j |\beta_j|)$$

Our Results

Laplace Prior Analysis

Prior mean $\beta = 0$			Prior mean $\beta = \hat{\beta}_{mle}$			Prior mean $\beta = \hat{\beta}_{post\ mean}$		
prior variance	training error (%)	testing error (%)	prior variance	training error (%)	testing error (%)	prior variance	training error (%)	testing error (%)
∞	28.00	29.82	∞	21.50	20.48	∞	23.00	26.81
10	22.00	21.99	10	21.50	20.48	10	21.50	19.88
100	21.50	19.56	$100\ \hat{\sigma}^2_{mle}$	21.50	20.48	$100\ \hat{\sigma}^2_p$	21.50	19.58
			$\hat{\sigma}^2_{mle}$	21.50	20.48	$\hat{\sigma}^2_p$	20.50	20.78

Gaussian Prior Analysis

	Full Model		Reduced Model	
	Training Error	Testing Error	Training Error	Testing Error
Pred Rule i)	.225	.193	.235	.214
Pred Rule ii)	..225	.193	.235	.217

Results from previous studies

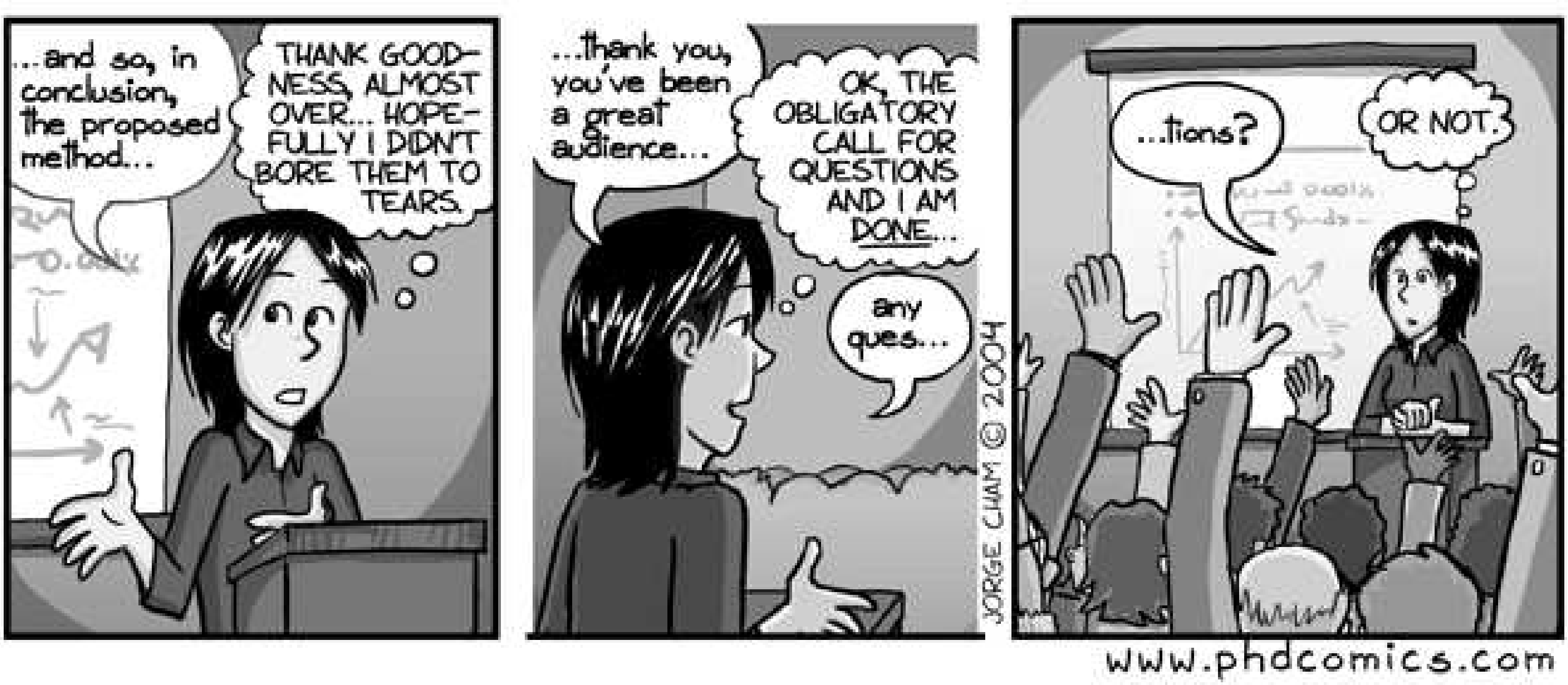
Algorithm	Max. Storage	Time (sec.) Train	Time (sec.) Test	Error Rate Train	Error Rate Test	Rank
Discrim	338	27.4	6.5	0.220	0.225	3
Quadisc	327	24.4	6.6	0.237	0.262	11
Logdisc	311	30.8	6.6	0.219	0.223	1
SMART	780	3762.0	*	0.177	0.232	4
ALLOC80	152	1374.1	*	0.288	0.301	21
k-NN	226	1.0	2.0	0.000	0.324	22
CASTLE	82	35.3	4.7	0.260	0.258	10
CART	144	29.6	0.8	0.227	0.255	9
IndCART	596	215.6	209.4	0.079	0.271	14
NewID	87	9.6	10.2	0.000	0.289	19
AC^2	373	4377.0	241.0	0.000	0.276	18
Baytree	68	10.4	0.3	0.008	0.271	14
NaiveBay	431	25.0	7.2	0.239	0.262	11
CN2	190	38.4	2.8	0.010	0.289	19
C4.5	61	11.5	0.9	0.131	0.270	13
ITrule	60	31.2	1.5	0.223	0.245	6
Cal5	137	236.7	0.1	0.232	0.250	8
Kohonen	62	1966.4	2.5	0.134	0.273	17
DIPOL92	52	35.8	0.8	0.220	0.224	2
Backprop	147	7171.0	0.1	0.198	0.248	7
RBF	179	4.8	0.1	0.218	0.243	5
LVQ	69	139.5	1.2	0.101	0.272	16
Default	*	*	*	0.350	0.350	23

Previous Results for this data (D. Michie, D.J. Spiegelhalter, and C.C. Taylor (1994), p158).

Conclusions

- Analysis using both types of prior families gives similar resutls.
- The error rates obtained using the two prediction rules in Gaussian prior analysis are close.
- For Laplacian prior, prior of mean of $\beta = 0$ and prior variance of 100 yield the best testing error percentage.
- Bayesian analysis methods give lower error rates than the previously studied learning methods.

Questions ?



Courtesy : http://www.phdcomics.com