John Fulgoni
Homework 1
COMS 4721
10 February 2015

Study Buddy: Mateus Braga – MA3382

John Fulgoni
Homework 1
10 February 2015

Problem 1 - Part 1

a) $\prod_{i=1}^{n} p(x_i \mid \pi)$

Bernoulli is $\pi^{x_i}(1-\pi)^{1-x_i}$

$$L(x_i \mid \pi) = \prod_{i=1}^{n} \pi^{x_i}(1-\pi)^{1-x_i} = \pi^{\sum_i x_i}(1-\pi)^{n-\sum_i x_i}$$

b) $$\hat{\pi}_{ML} = \ln\left(\pi^{\sum_i x_i}(1-\pi)^{n-\sum_i x_i}\right)$$

$$\nabla_\pi \left(\sum_{i=1}^{n} x_i \ln \pi + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-\pi) \right)$$

$$\frac{\sum_{i=1}^{n} x_i}{\pi} - \frac{n - \sum_i^{n} x_i}{1-\pi} = 0$$

mult by $\pi(1-\pi)$

$$(1-\pi)\sum_{i=1}^{n} x_i - \left(n - \sum_{i=1}^{n} x_i\right)\pi = 0$$

$$\sum_{i=1}^{n} x_i - \cancel{\pi \sum_{i=1}^{n} x_i} - n\pi + \cancel{\pi\sum_{i=1}^{n} x_i} = 0$$

$$\sum_{i=1}^{n} x_i = n\pi$$

$$\hat{\pi}_{ML} = \frac{\sum_{i=1}^{n} x_i}{n}$$

c) This makes sense because it is the average occurrence of x_i for all n. Since x is 0 or 1 they are equally likely

Part 2

a) poisson = $\frac{\lambda^x e^{-\lambda}}{x!} = p(x_i|\lambda)$

$L(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \sum_i \ln\left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}\right)$

b) $\sum_{i=1}^{n} x_i \ln(\lambda) + -\lambda \sum_{i=1}^{n} \ln(e) - \sum_i \ln(x!)$

$-\lambda \sum_{i=1}^{n} 1 = -\lambda n$

$\sum_{i=1}^{n} x_i \cdot \ln(\lambda) - \lambda n - \sum_{i=1}^{n} \ln(x!)$

$\nabla \lambda \quad \frac{\sum_{i=1}^{n} x_i}{\lambda} - n - 0 = 0$

$\frac{\sum_{i=1}^{n} x_i}{\lambda} = n$

$\hat{\lambda}_{mL} = \frac{\sum_{i=1}^{n} x_i}{n}$

c) This makes sense because the most likely value picked will be somewhere in the middle of the distribution.

Problem 2

a) Prior = $\frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$

proportional to likelihood · prior

$$p(\lambda | x_1, \ldots x_n) \propto \prod_{i=1}^{n} \left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \right) \cdot \frac{b^a}{\Gamma(a)} \lambda^{a-1} e^{-b\lambda}$$

$$\propto \lambda^{\sum_{i=1}^{n} x_i + a - 1} \cdot e^{-n\lambda - b\lambda}$$

$$\left(\sum x_i + a - 1\right) \ln \lambda + (-\lambda)(n+b) \qquad \nabla \lambda$$

$$\frac{\sum_{i=1}^{n} x_i + a - 1}{\lambda} - (n+b) = 0$$

$$\boxed{\lambda = \sum_{i=1}^{n} x_i + a - 1 \cdot (n+b)^{-1}}$$

This is proportional to a gamma distribution

b) Mean $= E[\lambda] = E\left[\frac{\sum_{i=1}^{n} x_i + a - 1}{(n+b)}\right]$

Mean of gamma $E[\lambda] = \frac{a}{b}$

$$= \frac{\sum^{n} x_i + a}{n+b}$$

MLE is of the same form as the mean.

This relates to the MLE value of the poisson distribution. As A and B approach 0, they will be nearly identical

Variance of gamma $E[2] = \frac{a}{b^2}$

The MLE relates to variance in the same way as the mean. If a and b approach 0, then the variance will approach the MLE

Part 1A:
wls
23.3589
-0.6704
1.0477
-0.0710
-5.8290
0.2495
2.7130

Each of the signs has a corresponding meaning for the legend:

Number of Cylinders (-) The more cylinders, the fewer MPG
Displacement (+) Higher displacement, the better MPG
Horsepower (-) Higher horsepower, the worse MPG
Weight (-) Higher weight, the worse MPG
Acceleration (+) Better acceleration, the better MPG
Model Year (+) The newer model year, the better MPG

Part 1B:

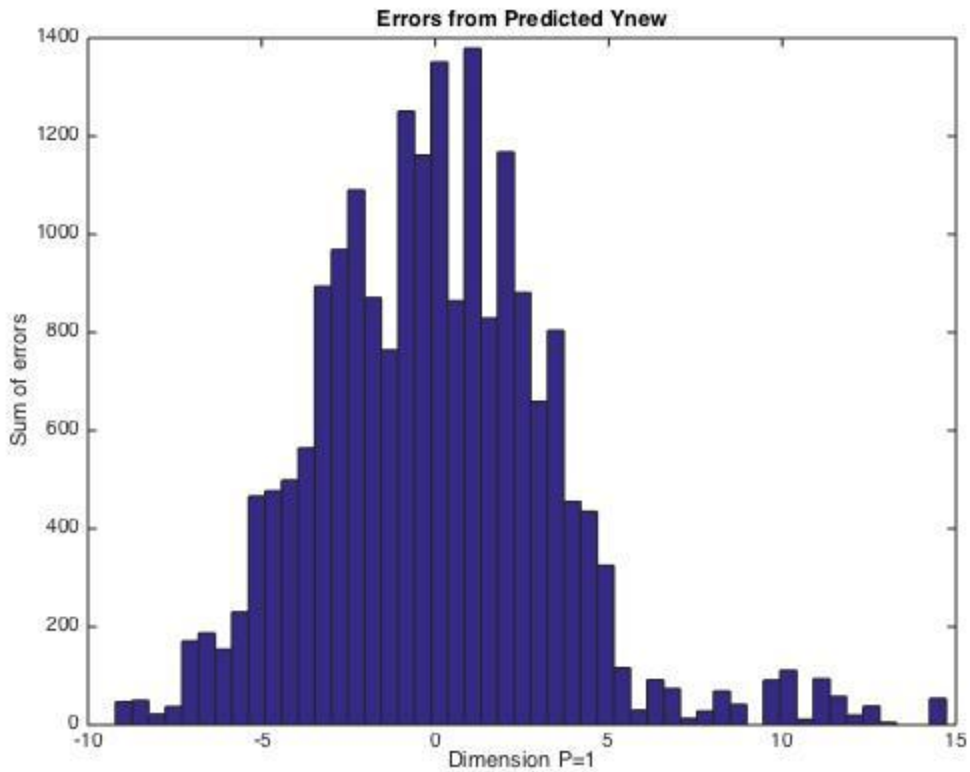
MAE_mean:
2.6771

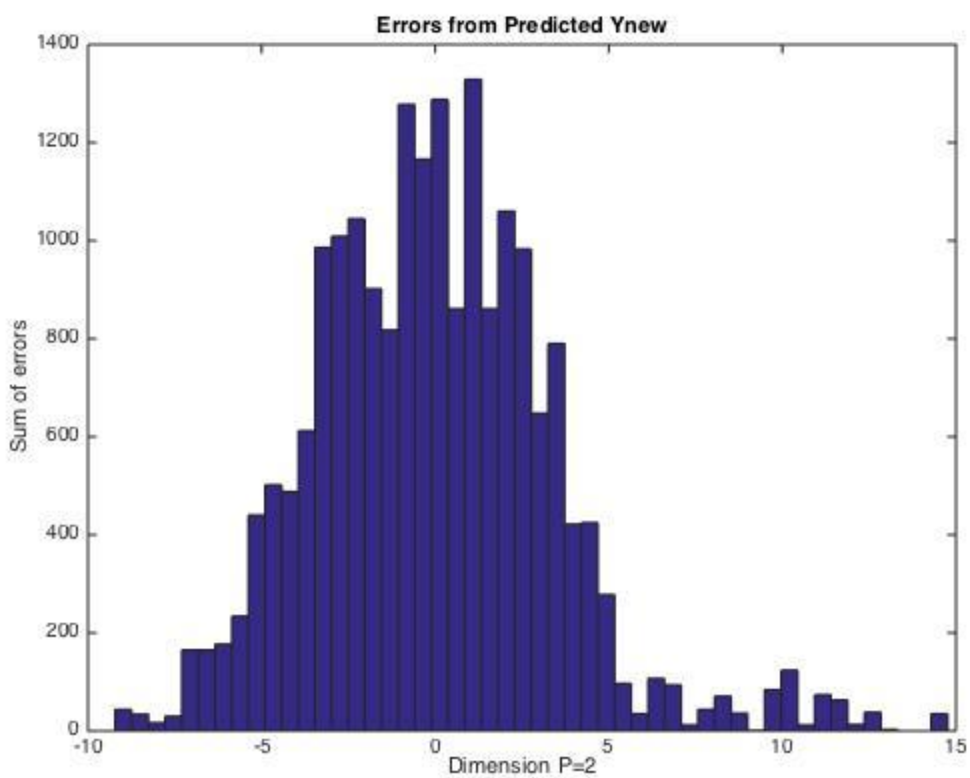
MAE_std:
0.4829

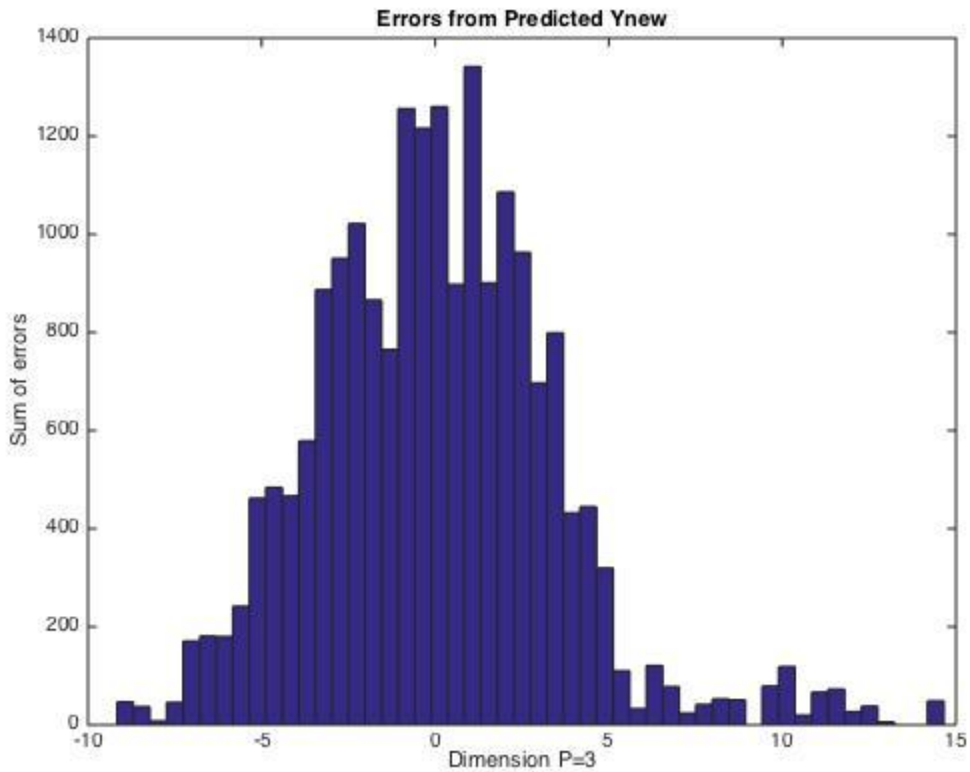
Part 2A:

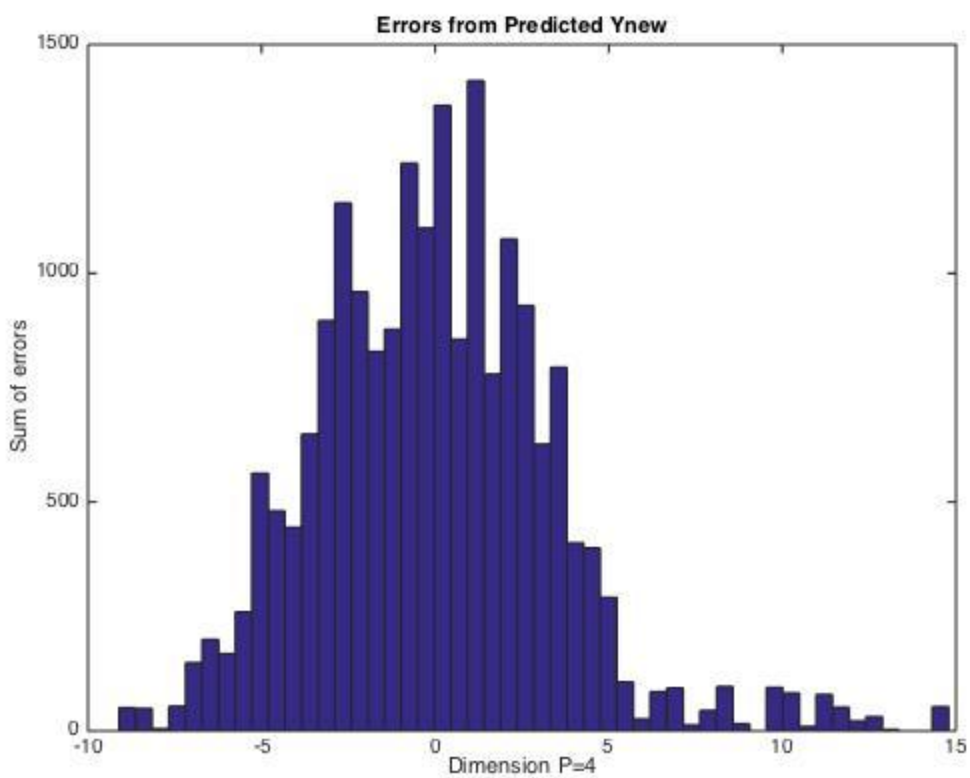
	R_means	R_stds
P=1	3.4226	0.69382
P=2	3.3819	0.66187
P=3	3.4195	0.6787
P=4	3.3976	0.66568

Part 2B:


Errors from Predicted Ynew
Sum of errors
Dimension P=1
1400
1200
1000
800
600
400
200
0
-10
-5
0
5
10
15


Errors from Predicted Ynew
Sum of errors
Dimension P=2
1400
1200
1000
800
600
400
200
0
-10
-5
0
5
10
15


Errors from Predicted Ynew
Sum of errors
Dimension P=3
1400
1200
1000
800
600
400
200
0
-10
-5
0
5
10
15


Errors from Predicted Ynew
Sum of errors
Dimension P=4
1500
1000
500
0
-10
-5
0
5
10
15

Part 2C:

The mean and variance for a Gaussian Distribution are calculated as:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\sigma = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

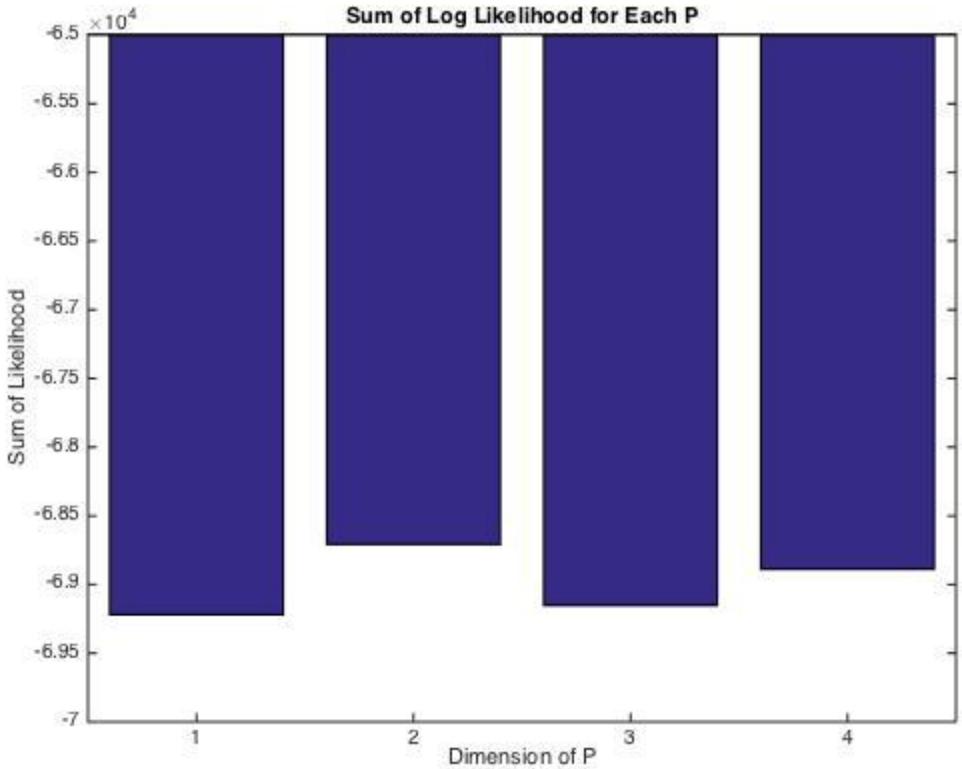
Using 20000 errors per Dimension of P

	Gaussian_Mean	Guassian_Variance
P=1	-0.0005546	12.195
P=2	-0.042047	11.873
P=3	0.0356	12.152
P=4	-0.026061	11.986

The log likelihood for a Guassian Distribution is calculated as:

$$\log\left(\frac{1}{\sigma\sqrt{2\pi}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}\right)$$

And we can sum this for all the errors for each p and compare them:



By looking at data from 2a, we can see that P=2 has the smallest RMSE as well as the smallest variance. By looking at the sum of the log-likelihood for each P, P=2 has the smallest log-likelihood for all of its errors. We can see that in general, the log-likelihood of the errors is correlated to the Root Mean Squared Error and its variance.

I would conclude by saying P=2 is the best choice for this model.