EECS E6720 Bayesian Models for ML.

HW 1, NAME: Qi Zhao. UNI: qz2316

$\boxed{\text{Problem 1}}$: I would suggest her to switch always.

Without loss of generality, we suppose she pick the first door D_1 and the host shows the second door D_2, otherwise, we could just change the number of the doors since they are identical and exchangeable. Then, we want to compare:

$$P(D_1=1 \mid pick\_D_1, show\_D_2) \quad \text{and} \quad P(D_3=1 \mid pick\_D_1, show\_D_2), \text{ where } D_i=1$$

means there is the prize behind that door.

Using Bayes Rule:

$$P(D_1=1 \mid Pick\_D_1, show\_D_2) = \frac{P(Pick\_D_1, show\_D_2 \mid D_1=1)\, P(D_1=1)}{P(Pick\_D_1, show\_D_2)} \Rightarrow \text{where } P(Pick\_D_1, show\_D_2 \cancel{\mid Pick\_D_1}$$

$$\mid D_1=1) = P(Pick\_D_1 \mid D_1=1)\cdot P(show\_D_2 \mid Pick\_D_1, D_1=1) = \frac{1}{3}\times\frac{1}{2}$$

$$\therefore P(D_1=1 \mid Pick\_D_1, show\_D_2) = \frac{\frac{1}{3}\times\frac{1}{2}\times\frac{1}{3}}{P(Pick\_D_1, show\_D_2)} \quad \because P(D_i=1)=\frac{1}{3}$$

With the same manner:

$$P(D_3=1 \mid Pick\_D_1, show\_D_2) = \frac{P(show\_D_2 \mid Pick\_D_1, D_3=1)\,P(D_3=1)\,P(Pick\_D_1 \mid D_3=1)}{P(Pick\_D_1, show\_D_2)}$$

$$= \frac{1\times\frac{1}{3}\times\frac{1}{3}}{P(Pick\_D_1, show\_D_2)} = 2P(D_1=1 \mid Pick\_D_1, show\_D_2),$$ so she should switch the door.

$\boxed{\text{Problem 2}}$:

For getting the proper prior, we should first check the likelihood:

$$P(\vec{X} \mid \pi) \propto \prod_{i=1}^{N} \pi_1^{X_{i1}} \pi_2^{X_{i2}} \cdots \pi_k^{X_{ik}} = \pi_1^{\sum_{i=1}^N X_{i1}} \pi_2^{\sum_{i=1}^N X_{i2}} \cdots \pi_k^{\sum_{i=1}^N X_{ik}},$$ where $\vec{X}$ is a Matrix, X_i is a k-dim vector $X_i = (X_{i1}, X_{i2}, \cdots X_{ik})^T$, based on this, we need a prior, whose parameters are all in the form of power. That is, Dirichlet dist. $\pi \sim Dirichlet(\alpha_1, \alpha_2, \cdots \alpha_k)$

$$\therefore P(\pi) \propto \pi_1^{\alpha_1-1} \pi_2^{\alpha_2-1} \cdots \pi_k^{\alpha_k-1}$$

$$\therefore \text{posterior}: P(\pi \mid \vec{X}) \propto P(\vec{X} \mid \pi)\, P(\pi) \propto \pi_1^{\alpha_1+\sum_{i=1}^N X_{i1}-1} \pi_2^{\alpha_2+\sum_{i=1}^N X_{i2}-1} \cdots \pi_k^{\alpha_k+\sum_{i=1}^N X_{ik}-1}$$

$$\therefore \pi \mid \vec{X} \sim Dirichlet(\alpha_1+\sum_{i=1}^N X_{i1}, \alpha_2+\sum_{i=1}^N X_{i2}, \cdots, \alpha_k+\sum_{i=1}^N X_{ik})$$

As we can see, the posterior's parameters are just the sum of prior's parameters and sample values (or to be specific, the number of obeservation for each group k).

Problem 3 : (a) We know $x_i \overset{i.i.d}{\sim}$ Poisso(λ) and $\lambda \sim$ Gamma(a, b),

that is, prior : $p(\lambda) \propto \lambda^{a-1} e^{-b\lambda}$

likelihood : $p(\vec{x}|\lambda) = \prod_{i=1}^{N} \lambda^{x_i} e^{-\lambda}/x_i! \ \propto_{\lambda} \lambda^{\sum_{i=1}^{N} x_i} e^{-N\lambda}$

$\therefore$ posterior with Bayes Rule :

$$p(\lambda|\vec{x}) \propto p(\vec{x}|\lambda)\, p(\lambda) \propto \lambda^{a+\sum_{i=1}^{N} x_i - 1} e^{-(b+N)\lambda}$$

$\therefore \lambda|\vec{x} \sim$ Gamma$(a+\sum_{i=1}^{N} x_i,\ b+N)$

(b). $$p(x^*|\vec{x}) = \int_0^\infty p(x^*|\lambda)\, p(\lambda|\vec{x})\, d\lambda = \int_0^\infty \lambda^{x^*} \frac{e^{-\lambda}}{x^*!} \frac{(b+N)^{a+\sum_{i=1}^{N} x_i}}{\Gamma(a+\sum_{i=1}^{N} x_i)} \lambda^{a+\sum_{i=1}^{N} x_i - 1} e^{-(b+N)\lambda}\, d\lambda$$

$$= \frac{(b+N)^{a+\sum_{i=1}^{N} x_i}}{x^*!\,\Gamma(a+\sum_{i=1}^{N} x_i)} \int_0^\infty \lambda^{a+x^*+\sum_{i=1}^{N} x_i - 1} e^{-(b+N+1)\lambda}\, d\lambda \qquad (*)$$

here since we know $\int_0^\infty \lambda^{a+x^*+\sum_{i=1}^{N} x_i - 1} e^{-(b+N+1)\lambda}\, d\lambda = \frac{\Gamma(a+\sum_{i=1}^{N} x_i + x^*)}{(b+N+1)^{a+\sum_{i=1}^{N} x_i + x^*}}$

$$\therefore (*) = \frac{(b+N)^{a+\sum_{i=1}^{N} x_i}}{\Gamma(x^*+1)\,\Gamma(a+\sum_{i=1}^{N} x_i)} \cdot \frac{\Gamma(a+\sum_{i=1}^{N} x_i + x^*)}{(b+N+1)^{a+\sum_{i=1}^{N} x_i + x^*}}$$

$$= \frac{(a+\sum_{i=1}^{N} x_i + x^* - 1)!}{x^*!\,(a+\sum_{i=1}^{N} x_i - 1)!} \cdot \frac{(b+N)^{a+\sum_{i=1}^{N} x_i}}{(b+N+1)^{a+\sum_{i=1}^{N} x_i}} \cdot \frac{1}{(b+N+1)^{x^*}}$$

$$= \binom{a+\sum_{i=1}^{N} x_i + x^* - 1}{a+\sum_{i=1}^{N} x_i - 1} \left(\frac{b+N}{b+N+1}\right)^{a+\sum_{i=1}^{N} x_i} \left(\frac{1}{b+N+1}\right)^{x^*}$$

$\therefore x^*|\vec{x} \sim$ Negative Binomial $(a+\sum_{i=1}^{N} x_i,\ b+N)$.

Problem 4 : To calculate the probability of Negative Binomial given x^*, I take log of it first and uses the formula here :

$$\ln \binom{a+\sum_{i=1}^{N} x_i + x^* - 1}{a+\sum_{i=1}^{N} x_i - 1} = \ln\Gamma(a+\sum_{i=1}^{N} x_i + x^*) - \ln\Gamma(x^*+1) - \ln\Gamma(a+\sum_{i=1}^{N} x_i)$$

The results of Problem 4:

b) :

Before showing the 2 X 2 table, one thing in my code should be clarified. That is, there are some emails can not be determined. The reason is the values of several variables of 27 test email are larger enough to let log(predict_probability) = -inf, meaning predict_probability is just 0 for both y = 1 or y = 0. If we can not decide them, we tend to regard them are spam since too much same words in them. For example, the word "address" appears 143 times in 309^{th} testing email, which seems very likely it is a spam email. Although according to the real testing label it is not a spam email indeed, and thus the prediction accuracy declined a little due to our suspicion, I still think assigning spam label to those testing emails with too much same words is reasonable.

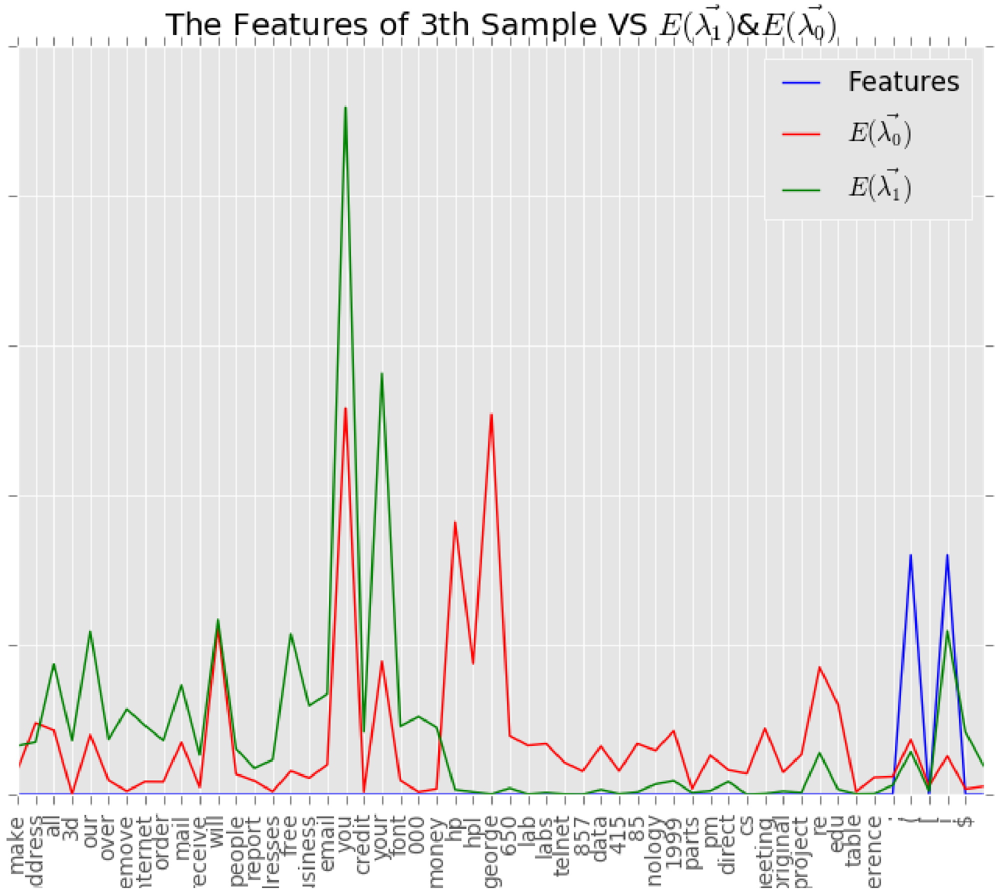
The 2 X 2 table is:

	Real Notspam	Real Spam
Predict Notspam	207	12
Predict Spam	47	170

So the prediction accuracy is around 82%.

c) The graphs and probabilities are shown below:

The first misclassified email is the 3^{rd} email in testing set.

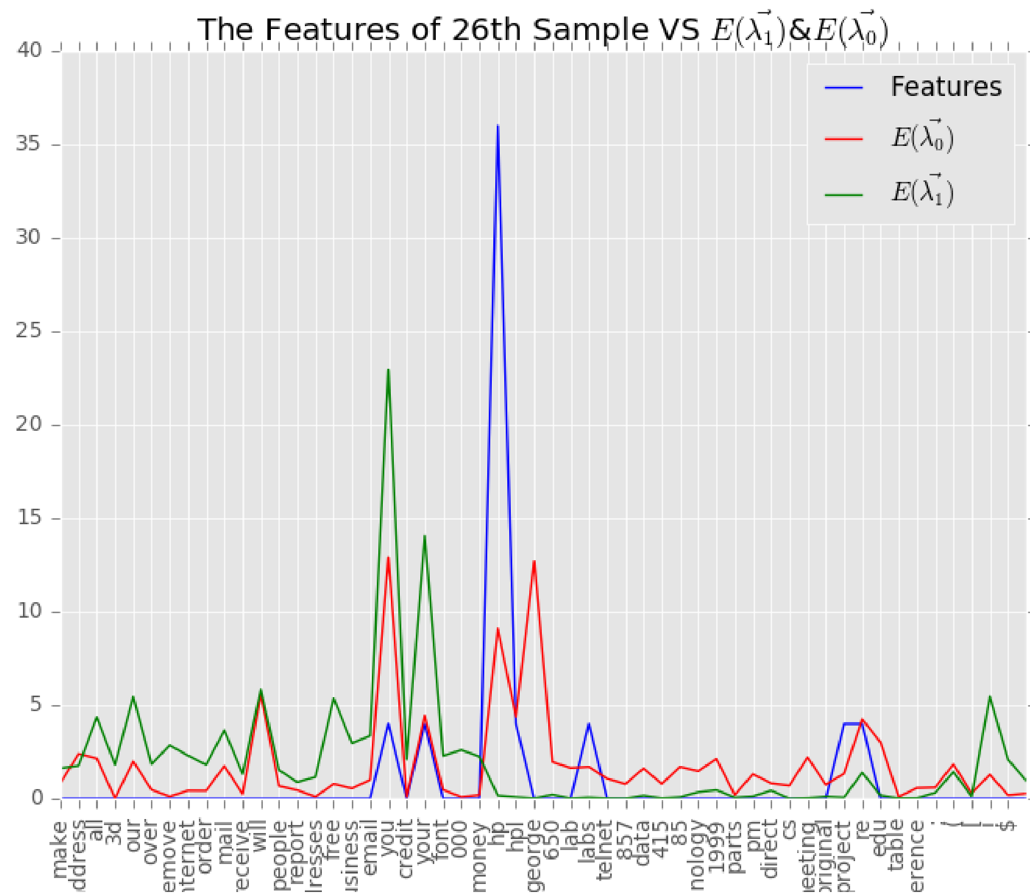


P(3rd Email is not Spam) = 0.928768807478

P(3rd Email is Spam) = 0.0712311925215

While the 3rd Email is spam email indeed.

The second misclassified email is the 26th email in testing set.

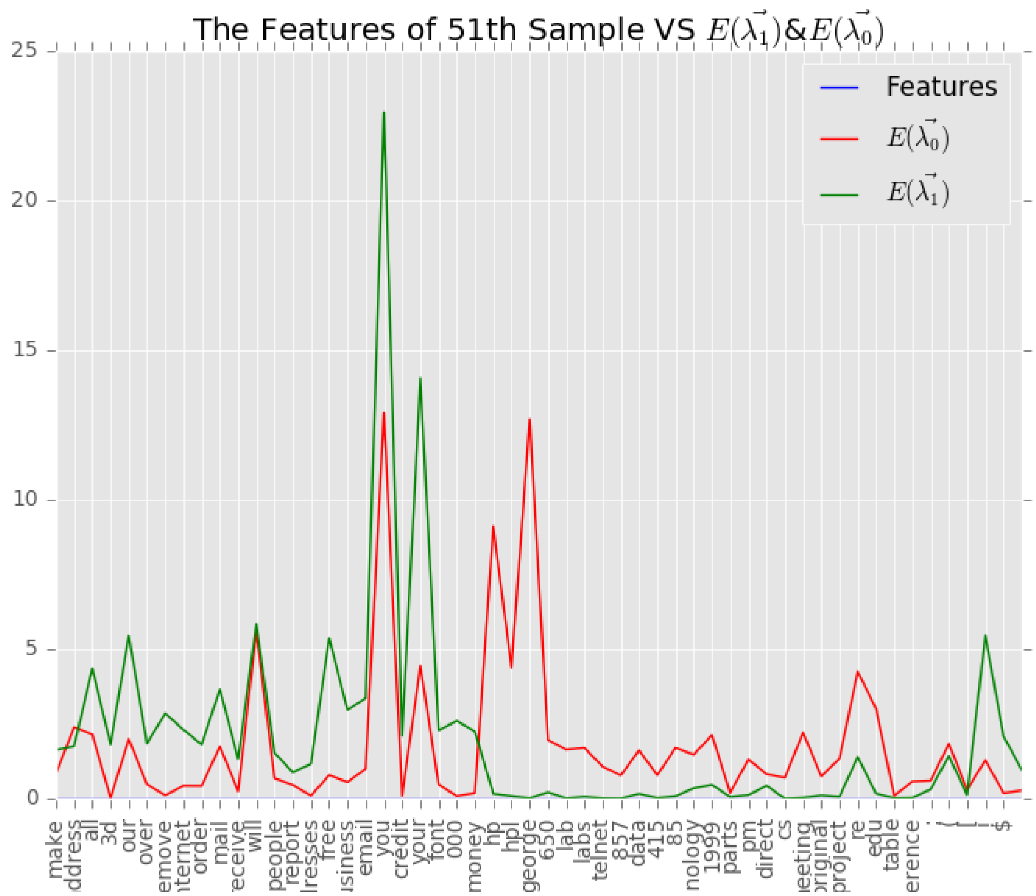


P(26th Email is not Spam) = 1.0

P(26th Email is Spam) = 1.13155686253e-84

While this is a Spam email indeed.

The third misclassified email is the 51th email in testing set.



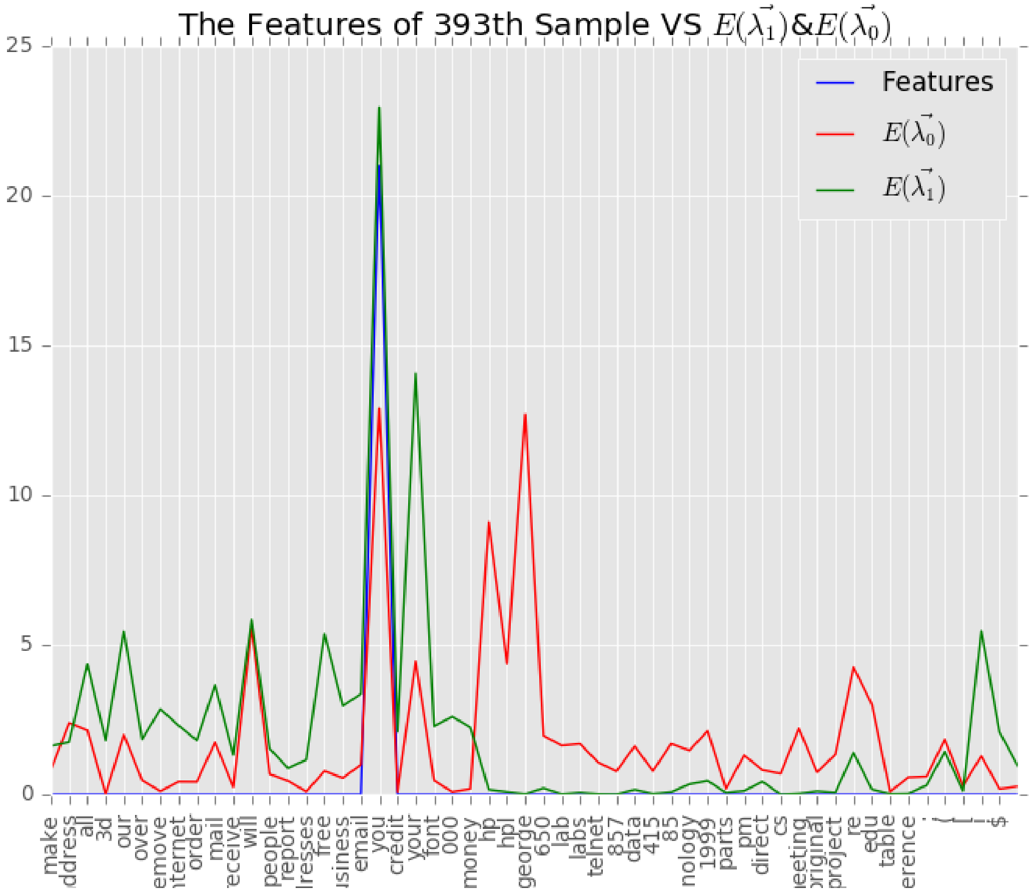
P(51th Email is not Spam) = 0.999994650979

P(51th Email is Spam) = 5.3490208913e-06

This email contains no words according to our 54 variables which we just believe this is not a spam email, while it is a spam email indeed.

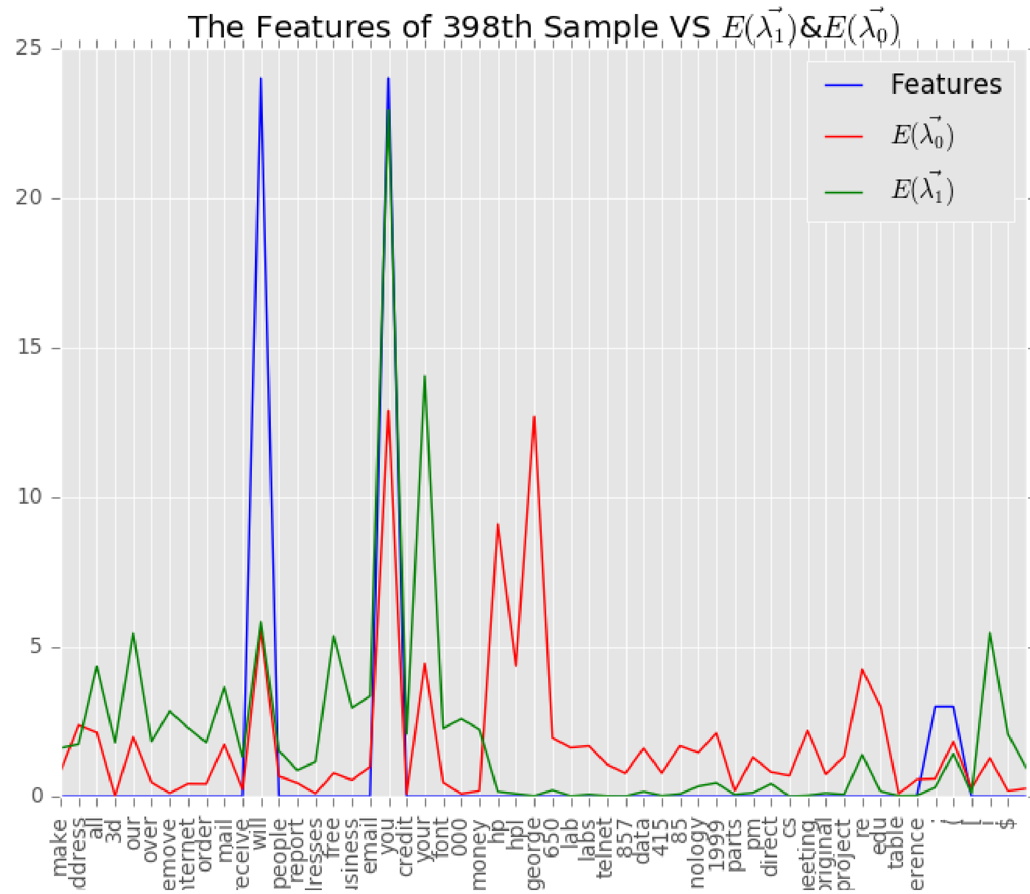
d) The graphs and probabilities are shown below:

The most ambiguous one is the 393rd email in testing set:



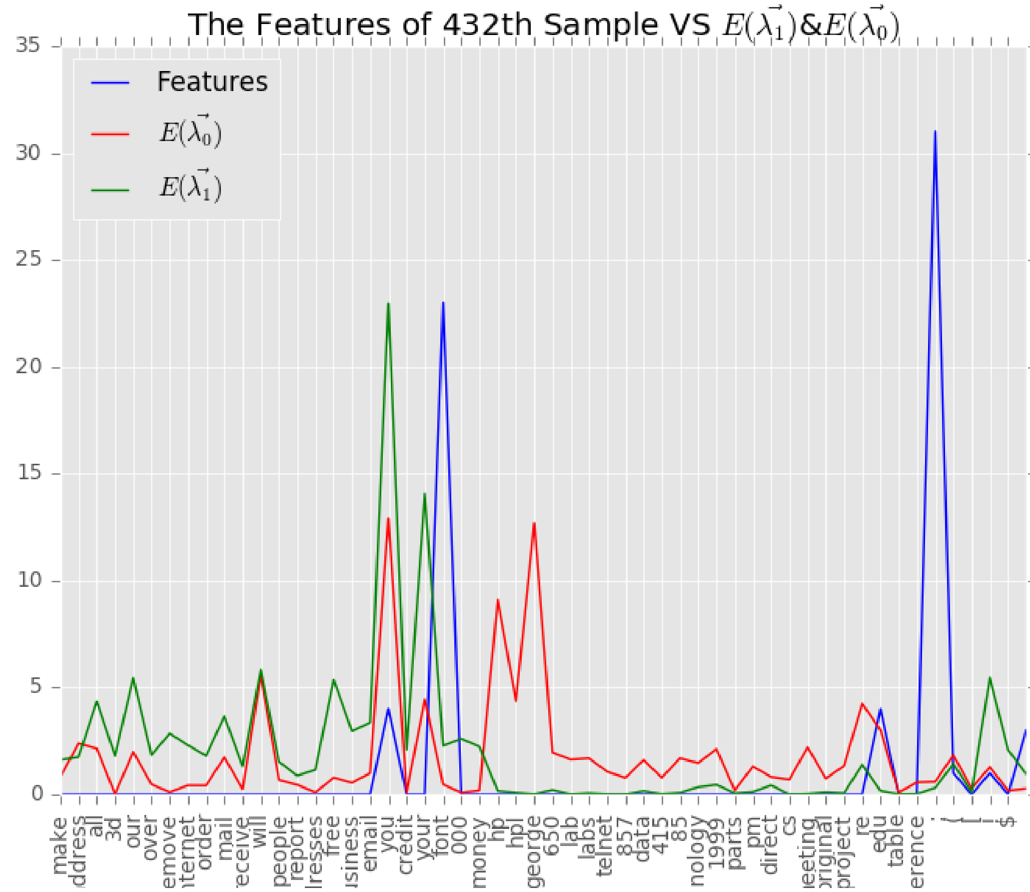
P(393th Email is not Spam) = 0.51062027739
P(393th Email is Spam) = 0.48937972261

The second most ambiguous one is 398th email in the testing set:



P(398th Email is not Spam) = 0.485782390087
P(398th Email is Spam) = 0.514217609913

The third most ambiguous one is the 432nd email in the testing set:



P(432th Email is not Spam) = 0.523072243961

P(432th Email is Spam) = 0.476927756039