P1. $y_i \sim Bern(\pi)$, $x_{i,d}|y_i \sim Pois(\lambda_{y_i,d})$ $\lambda_{y,d} \sim Gamma(2,1)$

(a) $\underset{\pi,\lambda_{y,D}}{\arg\max} \sum_{i=1}^{n} \ln p(y_i|\pi) + \sum_{d=1}^{D} (\ln p(\lambda_{0,d}) + \ln p(\lambda_{1,d}) + \sum_{i=1}^{n} \ln p(x_{i,d}|\lambda_{y_i,d})$

Only first part depends on π.

$\therefore \hat{\pi} = \underset{\pi}{\arg\max} \sum_{i=1}^{n} \ln p(y_i|\pi)$

$= \underset{\pi}{\arg\max} \sum_{i=1}^{n} \ln(\pi^{y_i}(1-\pi)^{1-y_i}) = \underset{\pi}{\arg\max} \sum_{i=1}^{n} (y_i \ln\pi + (1-y_i)\ln(1-\pi))$

By taking the derivative and setting to zero:

$$\sum_{i=1}^{n}\left(\frac{1}{\pi}y_i + \frac{y_i-1}{1-\pi}\right) = \sum_{i=1}^{n}\left(\frac{y_i}{\pi} + \frac{y_i-1}{1-\pi}\right) = 0 \Rightarrow \sum_{i=1}^{n}\left(\frac{y_i-\pi}{\pi(1-\pi)}\right) = 0$$

$$\hat{\pi} = \frac{\sum_{i=1}^{n} y_i}{n}$$

(b $\sum_{d=1}^{D}\{\ln\lambda_{0,d} e^{-\lambda_{0,d}} + \ln\lambda_{1,d} e^{-\lambda_{1,d}} + \sum_{i=1}^{n}[(1-y_i)\ln\frac{\lambda_{0,d}^{x_i} e^{-\lambda_{0,d}}}{x_i!} + \ln\frac{\lambda_{1,d}^{x_i} e^{-\lambda_{1,d}}}{x_i!} y_i]\}$

① MLE for λ_0

By taking derivative:

$$\left[\frac{1}{\lambda_{0,d}} - 1 + \sum_{i=1}^{n}\left(\frac{x_i}{\lambda_{0,d}} - 1\right)(1-y_i)\right] = 0$$

$$[1 - \lambda_{0,d} + \sum_{i=1}^{n}(x_{i,d} - \lambda_{0,d})(1-y_i)] = 0$$

$$[1-\lambda_{0,d} + \sum_{i=1}^{n}(1-y_i)x_{i,d} - \sum_{i=1}^{n}(1-y_i)\lambda_{0,d}] = 0$$

$$1-\lambda_{0,d} + \sum_{i=1}^{n}(1-y_i)x_{i,d} - \sum_{i=1}^{n}(1-y_i)\lambda_{0,d} = 0$$

$$\hat{\lambda_{0,d}} = \frac{1+\sum_{i=1}^{n}(1-y_i)x_{i,d}}{1+\sum_{i=1}^{n}(1-y_i)}$$

② MLE for λ_1

By taking derivative:

$$\left[\frac{1}{\lambda_{1,d}} - 1 + \sum_{i=1}^{n} y_i(x_{i,d} - \lambda_{1,d})\right] = 0$$

$$\frac{1}{\lambda_{1,d}} - 1 + \sum_{i=1}^{n} y_i x_{i,d} - \sum_{i=1}^{n} y_i \lambda_{1,d} = 0$$

$$\hat{\lambda_{1,d}} = \frac{1+\sum_{i=1}^{n} y_i x_{i,d}}{1+\sum_{i=1}^{n} y_i}$$

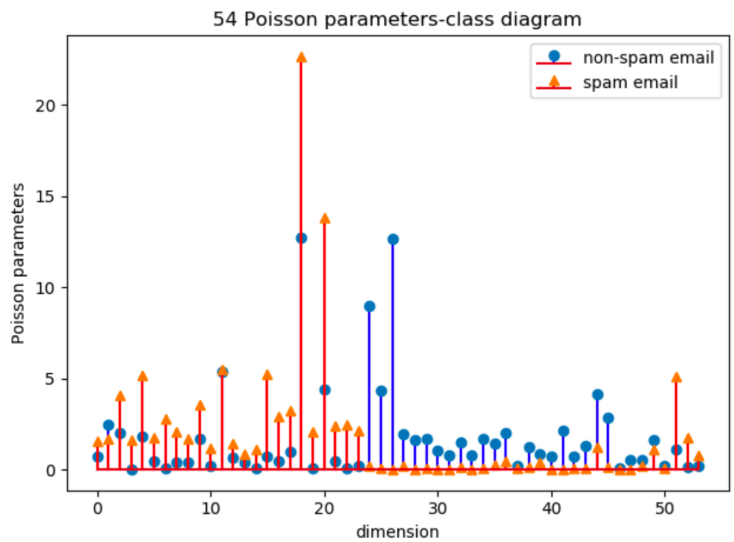
P2.

(a)

True\Pred	1	0
1	1712	101
0	491	2296

Accuracy = 87.13%

(b)

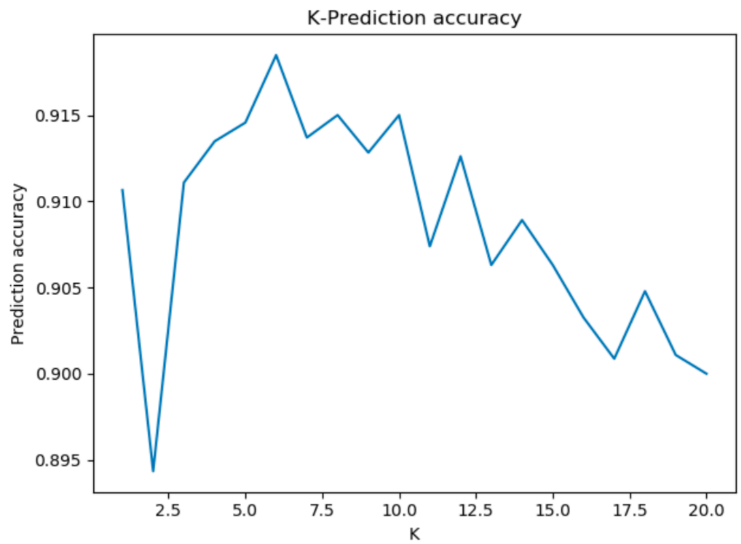


Dimension 16 : "free", The possion parameter for Spam email is much higher than that for non-spam email. The posion parameter for spam email is around 5.

Dimension 52 : " ! ", The poission parameter for spam email is higher than that for non-spam email. It's around 2 or 3.

These 2 dimensions' parameters are not the highest one.
It seems that "free" and "!" more likely shows in spam email.
This is in line with our common sense. Span emails always use "free" "!" to attract people.

(c) KNN's result.



P3.

a) shown as code.

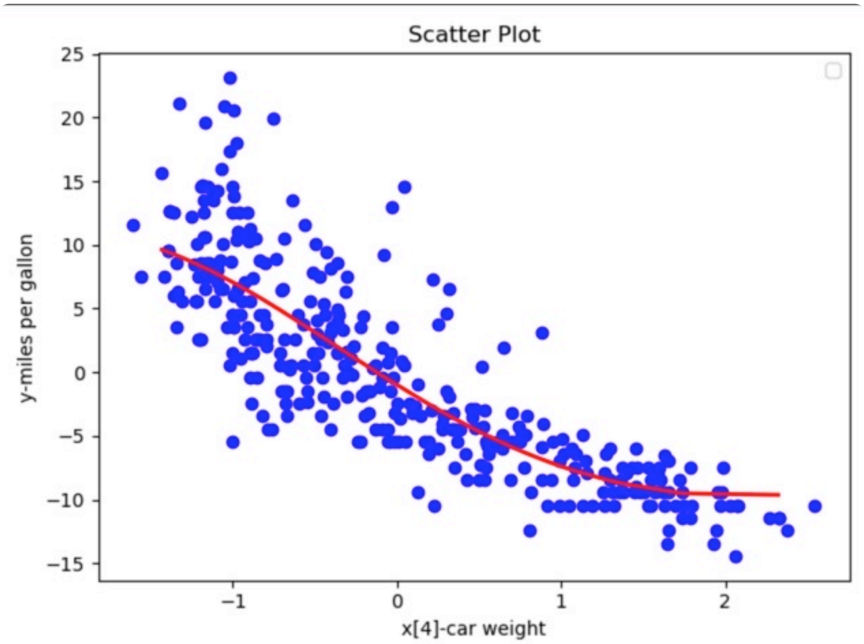
b)

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
5	1.9662770832011400	1.9331367430388900	1.9234218492955000	1.9221994122141700	1.9247710572433900	1.929214584821610	1.9346361805691700	1.9405852698053700	1.9468221527738300	1.9532147535652100
7	1.9201644559708700	1.9048781595333700	1.9080822266835800	1.915903742960170	1.9248062582061500	1.9337036635460000	1.942256089136370	1.9503824005884000	1.9580954559045700	1.9654404467105400
9	1.8976501853604700	1.902521011958540	1.9176498037931700	1.9325165811150100	1.9457018068985400	1.9572371755057400	1.9674056705298200	1.9764941762129700	1.9847432013773700	1.992343635707730
11	1.8905087802491300	1.9149830815209600	1.938851035693020	1.9579385936670300	1.973218217700310	1.9857666177104800	1.9963775483262600	2.0056056170174600	2.013837822849690	2.0213471533170600
13	1.895850359906790	1.9355881004200100	1.9645996488381400	1.9855043698573200	2.0013166575965500	2.013880822760730	2.024312725963830	2.0333090940990900	2.041319769316060	2.048643790039850
15	1.9096052208844500	1.959551103489170	1.9908059319304000	2.0119178086292600	2.027372618959860	2.039467464315050	2.0494656331464800	2.058107096465770	2.065847425140800	2.072978163729000

(c) The minimum RMSE is 1.8905 with $b=11$, $\sigma^2=0.1$. This result is better than what we did in homework 1. In homework 1. the RMSE is between around 2.65 to 2.85. In this method, the RMSE is around 1.89 to 2.07, which is lower than that in homework 1. In this case, Gaussian process is better.

The drawback is that GP's computation time. In practice it's hard to work with more than a few thousand points. The cost of GP computation is very high.

(d)


Scatter Plot
y-miles per gallon
x[4]-car weight
25
20
15
10
5
0
−5
−10
−15
−1
0
1
2