Q 2.1.1) log likelihood function of x given λ

$$P(X = k \mid \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k \in \{0, 1, 2, \ldots\}$$

$$\text{Likelihood} = \prod_{i=1}^{n} P(X = k_i \mid \lambda)$$

$$= \prod_{i=1}^{n} \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}$$

- Since all observations are independent.

Log likelihood:

$$\ln\left(\prod_{i=1}^{n} e^{-\lambda} \cdot \frac{\lambda^{k_i}}{k_i!}\right)$$

$$= \sum_{i=1}^{n} \ln\left(e^{-\lambda} \cdot \frac{\lambda^{k_i}}{k_i!}\right)$$

$$= \sum_{i=1}^{n} \left(\left(\ln e^{-\lambda}\right) + \left(\ln \frac{\lambda^{k_i}}{k_i!}\right)\right)$$

$$= \sum_{i=1}^{n} \left[(\ln e^{-\lambda}) - (\ln k_i!) + (\ln \lambda^{k_i}) \right]$$

$$= \sum_{i=1}^{n} \left[(-\lambda) - (\ln k_i!) + (k_i \ln \lambda) \right]$$

$$= -n\lambda - \sum_{i=1}^{n} (\ln k_i!) + \ln \lambda \sum_{i=1}^{n} (k_i)$$

Q2.1.2) MLE for λ

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}} \text{ (likelihood)}$$

$$= \underset{\lambda}{\operatorname{argmax}} \text{ (loglikelihood)}$$

Differentiating log likelihood wrt λ;

$$-n + \frac{1}{\lambda} \sum_{i=1}^{n} k_i = 0$$

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

Q2.1.3) MLE using observed x :

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

$$= \frac{1}{7}(4+5+3+5+6+9+3)$$

$$= \frac{1}{7}(35)$$

$$= \underline{\underline{5}}$$

Q2.2.1) $P(\lambda | \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \quad \lambda > 0$

Delays ressemble Gamma ~~[illegible]~~ distribution.

$\lambda \sim \Gamma(\alpha, \beta)$

$\text{mean} = \frac{\alpha}{\beta}$

$\text{mode} = \frac{\alpha - 1}{\beta} \quad \text{for } \alpha > 1$

prior distribution of $\lambda = \Gamma(\lambda | \alpha \beta)$

$P(x = k | \lambda) = \frac{\lambda^{k}}{k!} e^{-\lambda}$

$P(x=k|\lambda)$

$P(\lambda|\alpha\beta)$

$P(\lambda|x=k)$

$= \dfrac{P(x=k|\lambda)\ P(\lambda)}{P(x=k)}$

$= P(x=k|\lambda)\ P(\lambda)$

$\lambda^{\Sigma x_i+\alpha-1}\ e^{-\lambda(n+\beta)}$

$= e^{-\lambda n}\ \lambda^{\Sigma x_i}\ \lambda^{\alpha-1} \cdot e^{-\lambda n} \cdot e^{-\lambda\beta}$

$= \left(\lambda^{\Sigma x_i} \cdot e^{-\lambda\beta}\right) \cdot \left(\lambda^{\alpha-1} \cdot e^{-\lambda n}\right)$

$= \lambda^{\Sigma x_i} \cdot e^{-\lambda n} \left(\lambda^{\alpha-1} \cdot e^{-\beta\lambda}\right)$

$P(x=k|\lambda) = \dfrac{\lambda^k}{k!} e^{-\lambda}$

$$\frac{1}{\Gamma(\alpha)} \lambda \; e^{...} \cdot \frac{\lambda^k}{k!} e^{-\lambda}$$

$$= e^{-\beta\lambda - \lambda} \cdot \lambda^{\alpha-1+k} \left(\frac{\beta^\alpha}{k!\,\Gamma(\alpha)}\right)$$

prior of λ is $\Gamma(\lambda \mid \alpha, \beta)$

$\therefore$ posterior is $\rightarrow$

~~P(x)|λ)~~ =

$$P(\lambda \mid x) = \frac{P(x \mid \lambda)\, P(\lambda)}{P(x)}$$

$$\rightarrow \left(\lambda^n \cdot e^{-\lambda n}\right) \left(\lambda^{\alpha-1} \cdot e^{-\beta\lambda}\right)$$

$$P(\text{Data} \mid \text{model}) = P(x = k_i \mid \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$P(\text{model}) = P(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$\therefore$ P(model data) =

$$P(\lambda \mid x = k_i) = \frac{\frac{\lambda^k}{k!} e^{-\lambda}}{\int_0^\infty}$$

$2^3 = 8$ $\qquad 2^6 = 8 \times 8$

$= 2^3 \cdot 2^3$

Q2.3.1

$\eta = e^{-2\lambda}$

$X \sim \text{Poisson}(\lambda)$

$$P(X = k_i \mid \lambda) = \frac{\lambda^{k_i}}{k!} e^{-\lambda}$$

$$\hat{\eta} = e^{-2x}$$

$$\eta = e^{-2\lambda}$$

$$\cancel{\eta = (e^{-\lambda})^2}$$

$$e^{-2} \cdot e^{\lambda} = \eta$$

$$e^{\lambda} = \frac{\eta}{e^{-2}}$$

$$\lambda = \ln\left(\frac{\eta}{e^{-2}}\right)$$

$$\therefore \; P(x = ke \mid n) = \frac{\left(\ln\left(\frac{n}{e^{-2}}\right)\right)^{ke}}{k!} \quad \frac{e^{-2}}{n} = M$$

MLE for $n \rightarrow$

Taking log;

likelihood = ~~argmax~~ $\underset{n}{\text{argmax}}\; M$

$\therefore$ loglikelihood $= \log M$

$$\rightarrow ke \log\left(\log\left(\frac{n}{e^{-2}}\right)\right) - \log k! + \log\left(\frac{e^{-2}}{n}\right)$$

$\because$ ignore $(-\log k!)$

$$= ke \log\left(\log \frac{n}{e^{-2}}\right) + \log e^{-2} - \log n$$

$$= ke \log\left[\log n - \log e^{-2}\right] - \log n$$

ignore $\log e^{-2}$

$$= ke \log\log(n) - \log(n)$$

$$\cancel{= ke \log(\log n}$$

$$\therefore \frac{d}{dn} = \frac{ke}{n \log n} - \frac{1}{n} = \frac{1}{n}\left(\frac{ke}{\log n} - \log n\right)$$

$$= 0$$

Q2.3.1)

$\therefore$ ~~log n = kc~~

$$n = e^{-2\lambda}$$

$$\therefore \quad n^{1/2} = e^{-\lambda}$$

$$n^{-1/2} = e^{\lambda}$$

$$\lambda = \log n^{-1/2}$$

$$= \frac{-1}{2}\log n$$

$$P(x=k_i \mid \lambda) = \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}$$

$$\therefore P(x=k_i \mid n) = \frac{(-1/2 \log n)^{k_i}}{k_i!} e^{\frac{-1}{2}\log n}$$

$$= \frac{(-1/2 \log n)^{k_i}}{k_i!} n^{-1/2}$$

$$\text{likelihood} = \underset{n}{\text{argmax}} \frac{(-1/2 \log n)^{k_i}}{k_i!} n^{-1/2}$$

$$\text{log likelihood} = \underset{n}{\text{argmax}} \log\left(\frac{-1/2 \log n}{k_i!}\right)^{k_i} + \log n$$

$$= \underset{n}{\operatorname{argmax}} \left[k_i \log\left(\frac{-1/2 \log n}{k_i!}\right) + -\frac{1}{2}\log n \right]$$

$\rightarrow 1/2 \log n = m$

$$= \underset{n}{\operatorname{argmax}} \left[k_i \log(-m) - k_i \log k_i! \cancel{+\frac{1}{2}} + \bar{m} \right]$$

$\cancel{m = \frac{1}{2}\log n}$

Differentiating wrt n;

$\cancel{\frac{dm}{dn} = \frac{1}{2}\frac{1}{n}}$

$$\frac{d}{dn}\left(k_i \log(-1/2 \log n) \cancel{+} -1/2 \log n\right)$$

$$= \frac{k_i}{-1/2 \log n} \cdot -\frac{1}{2}\frac{1}{n} \cancel{+} -\frac{1}{2n}$$

$$= \frac{1}{2n}\left(\cancel{k_i} - 1 - \frac{k_i}{-1/2 \log n}\right)$$

$$\therefore \quad 1/2 \log n = -k_i = x$$

$$\log n = \cancel{\quad} -2x$$

$$n = e^{-2x}$$

Q3.1.1) $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$

$\epsilon_1 \dots \epsilon_n$ are independent Gaussians.

$$y_i = w^T x_i + \epsilon_i$$

Now; log likelihood calculation :

$$\sum_{i=1}^{N} \log \left[\left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp\left(\frac{-1}{2\sigma_i^2} (y_i - w^T x_i)^2 \right) \right]$$

$$= \sum_{i=1}^{N} \left[\log \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} - \frac{1}{2\sigma_i^2} (y_i - w^T x_i)^2 \right]$$

$$= \sum_{i=1}^{N} \left[\underbrace{\begin{matrix} \text{some } \sigma_i \\ \text{term} \end{matrix}}_{\text{constant}} \quad -\frac{1}{2\sigma_i^2} (y_i - w^T x_i)^2 \right]$$

$$\text{minimize} \sum_{i=1}^{N} \frac{-1}{2\sigma_i^2} (y_i - w^T x_i)^2$$

$$s_i = \frac{-1}{2\sigma_i}$$

$$\frac{\partial}{\partial w} \sum s_i (y_i - w x_i)^2 =$$

$$2 s_i \sum -x_i (y_i - w x_i)$$

$$\Rightarrow \sum 2 s_i x_i (y_i - w x_i) = 0$$

$$\sum 2 s_i x_i y_i - \sum 2 s_i x_i^2 w = 0$$

$$\therefore \sum x_i y_i = \sum w x_i^2$$

$$w = \frac{\sum x_i y_i}{\sum x_i^2}$$

Q3.2) $p(\varepsilon_i) = \frac{1}{2b} \exp\left(-\frac{|\varepsilon_i|}{b}\right) = \frac{1}{2b} e^{-\frac{|\varepsilon_i|}{b}}$

$y(x) = w^T x + \varepsilon$

$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \log p(D \mid \theta)$

$\Rightarrow$ loglikelihood:

$$\sum_{i=1}^{N} \log\left[\frac{1}{2b} e^{-\frac{|\varepsilon_i|}{b}}\right]$$

$$\Rightarrow \sum_{i=1}^{N} \log\left[\frac{1}{2b} e^{-\frac{|y_i - wx_i|}{b}}\right]$$

$$= \sum_{i=1}^{n} \left[\log(1/2b) * -|y_i - wx_i|\right]$$

$$= \frac{1}{b} \sum_{i=1}^{n} |x_i w - y_i| + \underbrace{n \log 1/2b}_{\text{constant}}$$

maximize loglikelihood $= \cancel{\text{maximize}}$ $\frac{1}{b}\sum_{i=1}^{n}(x_i w - y_i)$

$$= \text{maximize} \sum_{i=1}^{n} |x_i w - y|$$

1277

Q4.1.1 $r_i = y_i - (w^T x_i + b)$

$\|x\|_0$ = no. of non zero x entries

$R = Y - W^T X - B$

Time complexity = $O(\|X\|_0)$

Q4.1.2

$B_{new} = \frac{R + B_{OLD}}{N}$

Time complexity = $O(n)$ ~~=~~ $\to$ $O(\|X\|_0)$

where $B \Rightarrow (N \times 1)$

~~Q4.1.3 $R_{new} = R_{OLD} - B_{NEW} + B_{OLD}$~~

~~$O(n)$~~

~~Q4.1.4 $C_k = 2 [R + W_k X_k] X_k$~~

~~vector X_k~~

~~kth element of W~~

~~Q4.1.5~~

Q4.1.3

$$R_{New} = R_{OLD} - B_{New} + B_{OLD}$$

$$O(n)$$

Q4.1.4 $C_K = 2\left[R + W_K X_K\right] X_K$

$$O(Z_K)$$

Q4.1.5 $R_{NEW} = R_{OLD} + W_{K_{OLD}}^T X_K - W_{K_{NEW}}^T X_K$

$$O(Z_K)$$

Q4.1.6 1. $O\|X\|_0$

2. $O(n)$ $\rightarrow$ $(n \times d)$ matrix considered

3. $O(n)$

~~4.~~

4. $(d)(Z_K)$

5. $(d)(1)$

6. $(d)(Z_K)$

$\therefore$ Total time $= O\|X\|_0 + O(n) + O(d) + O(dZ_K)$

Q4.3 The λ to precision-recall relation:

As shown in graphs;

precision increases as λ increases

Recall decreases as λ increases

Changing σ to 10 →

Precision decreased;

recall remains same

→ To achieve better precision & recall;

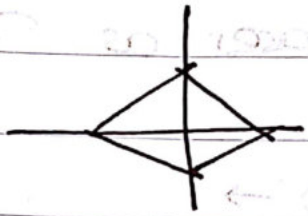
As seen through results; the λ should be increased to increase precision.

There will be a point after which recall starts falling. Stop increasing λ at this point...

Q 3-3] Laplace prefered over Gaussian -

1) laplace has thinner tail.

2) laplace - solutions intersect the



curve at

edges; which helps eliminate unneces features

3) Sparser solutions --

~~4) The error is~~

Q1. X_1 $[0, 1]$

X_2 $[0, 1]$

$X \rightarrow \max(X_1, 2X_2)$

$$f_{X_1} = \begin{cases} 1 & 0 < x_1 < 1 \\ 0 & \text{else} \end{cases}$$

$$f_{X_2} = \begin{cases} 1 & 0 < x_2 < 1 \\ 0 & \text{else} \end{cases}$$

$X_3 = 2X_2$

$$f_{X_3} = \begin{cases} 1/2 & 0 < x_3 < 2 \\ 0 & \text{else} \end{cases}$$

$$F_{x_1} = \begin{cases} 0 & x_1 < 0 \\ x & 0 < x_1 < 1 \\ 1 & x_1 \geq 1 \end{cases}$$

$$F_{x_3} = \begin{cases} 0 & x < 0 \\ x/2 & 0 < x_3 < 2 \\ 1 & x_3 \geq 2 \end{cases}$$

$$f_X = f_{X_1} F_{X_3} + f_{X_3} F_{X_1}$$

$$f_x = \begin{cases} 0 & x < 0 \\ 1/2 & 1 < x < 2 \\ \cancel{3/4}\, x & 0 < x < 1 \end{cases}$$

$$E(x) = \int_0^1 \frac{\cancel{3}}{\cancel{4}} x^2\,dx + \int_1^2 \frac{x}{2}\,dx$$

$$= \left[\frac{x^3}{3}\right]_0^1 + \left[\frac{x^2}{4}\right]_1^2$$

$$= \frac{1}{3} + \frac{3}{4}$$

$$= \frac{13}{12}$$

$(13/12)^2$

$$\text{var}(x) = E(x^2) - (E(x))^2$$

$$= \int_0^1 x^{\cancel{3}}\,dx + \int_1^2 \frac{x^2}{2}\,dx$$

$$= \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{x^3}{6}\right]_1^2$$

$$= \frac{1}{4} + \frac{8}{6} - \frac{1}{6} = \frac{1}{4} + \frac{7}{6}$$

$$= \frac{6 + 28}{24}$$

$$= \frac{34}{24}$$

$$= \frac{17}{12}$$

- $Cov(X Y) = E(X Y) - E(X) E(Y)$

→ $\cancel{Cov(X, X_1)} = \cancel{E(X X_1)} - \cancel{E(X)} E(X_1)$

$\cancel{13/12}$ $\cancel{17/12}$

$\cancel{E(X X_1)} = \cancel{\iint X X_1}$

Q4.3 The λ to precision - recall relation :
As shown in graphs ;
precision increases as λ increases
Recall decreases as λ increases

Changing σ to $10 \rightarrow$
Precision decreased ;
recall remains same

→ To achieve better precision & recall;
As seen through results ; the λ should be increased to increase precision.
There will be a point after which recall starts falling. Stop increasing λ at this point...

→ Smaller λ takes more iterations to converge.

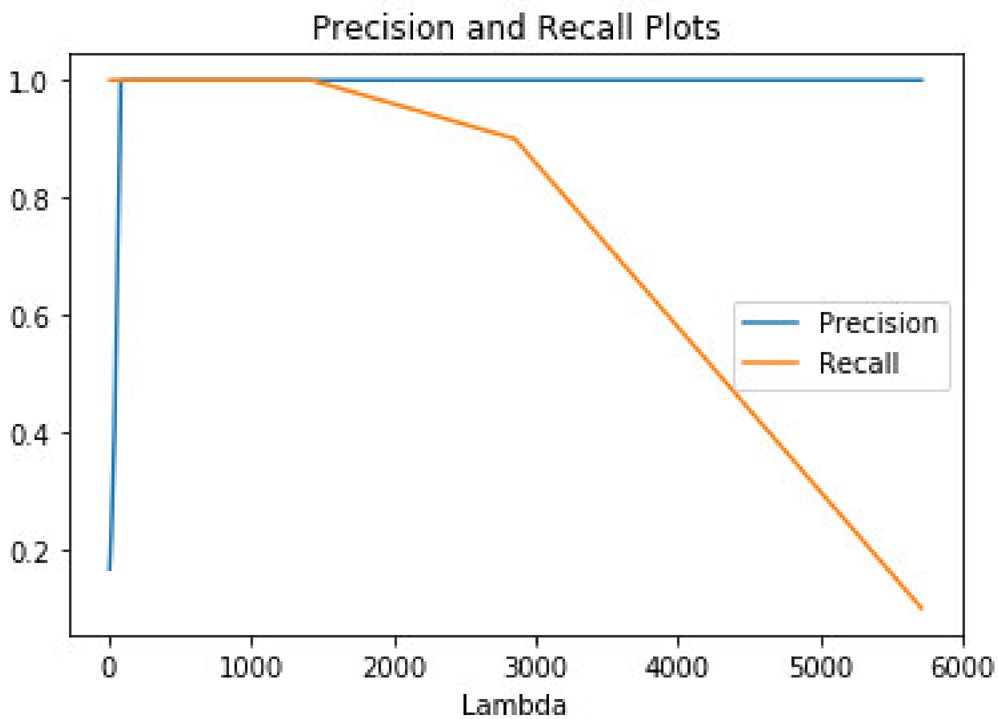
- Priors added to end of PDF

Q.4.3

1.

Starting from lambda max and finding precision recall for 10 iterartions, the following plot was made:

```
Precision [1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.4, 0.22727272727272727,
0.16666666666666666]
Recall [0.1, 0.9, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]
```



2.

Changing sigma to 10:

Lambda value of 713.78 was used

This gave the following result:

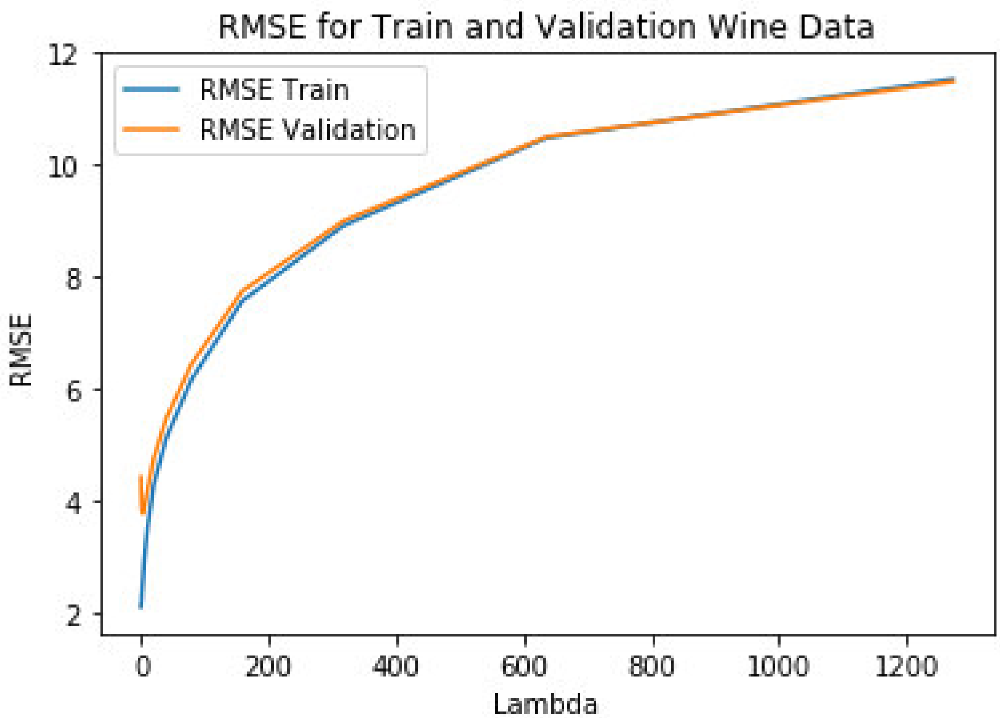
```
('Non zeros :', 10)
('Precision is : ', 0.15873015873015872)
('Recall is: ', 1.0)
```

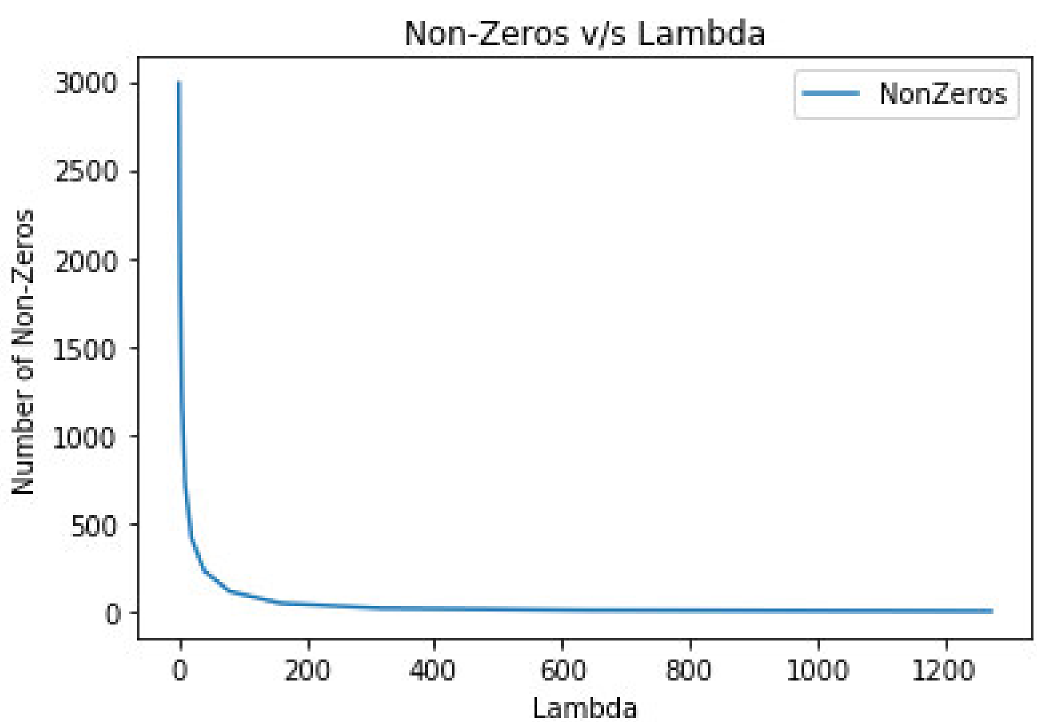
Precision was less but recall remained same.

Please refer the last page of handwritten notes for the explanations of 4.3

4.4

1.


RMSE for Train and Validation Wine Data
RMSE Train
RMSE Validation
RMSE
Lambda
12
10
8
6
4
2
0
200
400
600
800
1000
1200


Non-Zeros v/s Lambda
NonZeros
Number of Non-Zeros
Lambda
3000
2500
2000
1500
1000
500
0
0
200
400
600
800
1000
1200

2.

As seen from the RMSE plot,

RMSE on validation decreases till 3.79 after which it increases. The value of **lambda chosen = 2.48**

This achieves the best validation performance

Features with largest weights:

```
['spearmint\n',
 'stars\n',
 'big\n',
 'lifesaver\n',
 'ageability\n',
 'lemony\n',
 'sweet black\n',
 'nearly\n',
 'truly\n',
 'acidity provides\n']

Features with the smallest weights:

['earns\n',
 'high\n',
 'cherry berry\n',
 'soft\n',
 'sparkler\n',
 'liqueur\n',
 'cuts\n',
 'semillon\n',
 'banana\n',
 'brightened\n']
```

Words like “Lifesaver” and “big” contribute more the score and have higher weights than words like “soft” and “banana” which possibly people don’t like in wines. The weights therefore make some sense intuitively

3.

RMSE on Test Data from Kaggle : 2.02698