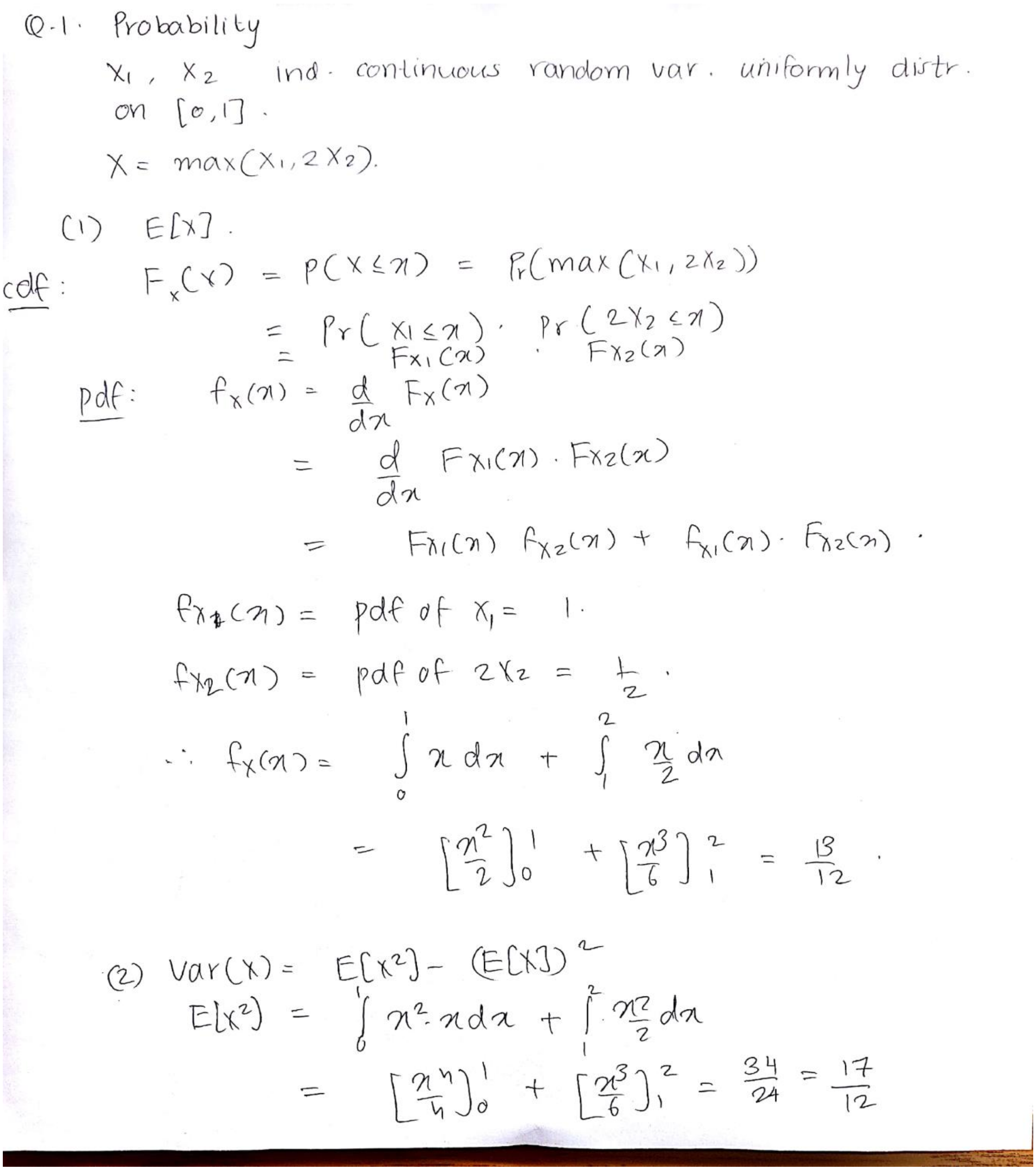
ML Assignment 1

Mihir Chakradeo 111462188



Q.1. Probability

X_1, X_2 ind. continuous random var. uniformly distr. on $[0,1]$.

$X = \max(X_1, 2X_2)$.

(1) $E[X]$.

cdf: $F_X(X) = P(X \leq x) = Pr(\max(X_1, 2X_2))$

$= Pr(X_1 \leq x) \cdot Pr(2X_2 \leq x)$

$= F_{X_1}(x) \cdot F_{X_2}(x)$

pdf: $f_X(x) = \frac{d}{dx} F_X(x)$

$= \frac{d}{dx} F_{X_1}(x) \cdot F_{X_2}(x)$

$= F_{X_1}(x) f_{X_2}(x) + f_{X_1}(x) \cdot F_{X_2}(x)$.

$f_{X_1}(x) =$ pdf of $X_1 = 1$.

$f_{X_2}(x) =$ pdf of $2X_2 = \frac{1}{2}$.

$\therefore f_X(x) = \int_0^1 x \, dx + \int_1^2 \frac{x}{2} \, dx$

$= \left[\frac{x^2}{2}\right]_0^1 + \left[\frac{x^3}{6}\right]_1^2 = \frac{13}{12}$.

(2) $Var(X) = E[X^2] - (E[X])^2$

$E[X^2] = \int_0^1 x^2 \cdot x \, dx + \int_1^2 \frac{x^2}{2} \, dx$

$= \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{x^3}{6}\right]_1^2 = \frac{34}{24} = \frac{17}{12}$

Q.2. Parameter Estimation.

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

(2.1) Give log-likelihood fn. of X given λ

① $L(X_1 \ldots X_n \mid \lambda) = P(X_1\mid\lambda) . P(X_2\mid\lambda) \ldots P(X_n\mid\lambda)$

$$= \prod_{i=1}^{n} P(X_i \mid \lambda)$$

Taking log likelihood

$$= \text{Log}\left(\prod_{i=1}^{n} P(X_i\mid\lambda)\right) = \text{Log}\left(\prod_{1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right)$$

$$= \sum_{i=1}^{n} \text{Log}\left(\frac{\lambda^{X_i} e^{-\lambda}}{X_i!}\right)$$

$$= \sum_{i=1}^{n} \left[\text{Log}(\lambda^{X_i} e^{-\lambda}) - \text{Log}\, X_i!\right]$$

$$\boxed{L(\lambda) = \text{Log}\,\lambda \cdot \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \text{Log}(X_i!)}$$

② Compute MLE of λ in general case.

~~Take derivative~~ $L(\lambda) \underset{\substack{\text{arg}\\ \text{max}\\ \lambda}}{=} \text{Log}\,\lambda \cdot \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \text{Log}(X_i!)$

Take derivative wrt. λ and set it to 0.

i.e $\frac{\partial}{\partial\lambda} L(\lambda) = 0. \Rightarrow \frac{\partial}{\partial\lambda}\left(\text{Log}\,\lambda \cdot \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \text{Log}(X_i!)\right) = 0.$

$$\therefore \frac{1}{\lambda} \sum_{i=1}^{n} X_i - n - 0 = 0.$$

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

③ Compute MLE for λ using the observed X.

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

$$\therefore \quad \lambda = \frac{1}{7} \times [4+5+3+5+6+9+3] = \frac{35}{7} = 5.$$

(2.2) MAP. $p(\lambda|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{\alpha-1} e^{-\beta\lambda}, \quad \lambda > 0$

① compute posterior distribution over λ.

$$P(\lambda | X_1 \ldots X_n) = P(X_1 \ldots X_n | \lambda) . P(\lambda) \quad --- ①$$

$$P(X_1 \ldots X_n | \lambda) = \prod_{i=1}^{n} P(X_i|\lambda)$$

we already have calculate $\prod_{i=1}^{n} P(X_i|\lambda)$ in (2.1.1)

$$\therefore P(X_1 \ldots X_n | \lambda) = \log\lambda \cdot \sum_{i=1}^{n} X_i - n\lambda - \sum_{i=1}^{n} \log(X_i!)$$

$$= \frac{\lambda^{\sum_{i=1}^{n} X_i} e^{-n\lambda}}{\prod_{i=1}^{n} (X_i!)}$$

Put in equation ①.

$$P(\lambda | X_1 \ldots X_n) = \frac{\lambda^{\sum_{i=1}^{n} X_i} e^{-n\lambda}}{\prod_{i=1}^{n} X_i!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \lambda^{\alpha-1} \cdot e^{-\beta\lambda}$$

$$\boxed{P(\lambda | X_1 \ldots X_n) = e^{-\lambda(n+\beta)} \cdot \lambda^{\sum_{i}^{n} X_i + \alpha - 1} \cdot \frac{\beta^\alpha}{\prod_{i=1}^{n} X_i! \cdot \Gamma(\alpha)}}$$

(2-2)

② Derive an analytic expression for the maximum a posterior (MAP) estimate of λ.

$$\lambda_{MAP} = \underset{\lambda}{\text{argmax}} \; \text{Log}\left(e^{-\lambda(n+\beta)} \cdot \lambda^{\sum_{i=1}^{n} X_i + \alpha - 1} \frac{\cancel{3} \cdot \beta^{\alpha}}{\prod_{i=1}^{n} X_i! \, \Gamma(\alpha)} \right)$$

~~Take derivative~~ $= \underset{\lambda}{\text{argmax}} \Big[-\lambda(n+\beta) + \left(\sum_{i=1}^{n} X_i + \alpha - 1\right) \text{Log}\,\lambda$

$$+ \text{Log}\left(\frac{\beta^{\alpha}}{\prod_{i=1}^{n} X_i! \, \Gamma(\alpha)} \right) \Big]$$

Take derivative wrt λ and set it to 0.

$$\therefore \quad 0 = -(n+\beta) + \frac{1}{\lambda} \sum_{i=1}^{n} X_i + \alpha - 1 .$$

$$\therefore \quad \boxed{\lambda = \frac{\sum_{i=1}^{n} X_i + \alpha - 1}{n+\beta}}$$

$X \sim \text{poisson}(\lambda)$, $\eta = e^{-2\lambda}$

(2-3)

① $\hat{\eta} = e^{-2X}$. Show that $\hat{\eta}$ is the MLE of η.

We have, $P(X=k|\lambda) = \dfrac{\lambda^{k} e^{-\lambda}}{k!}$ {poisson's distr.}

Also, $\eta = e^{-2\lambda} \Rightarrow \lambda = \dfrac{-\text{Log}\,\eta}{2}$.

Subs. value of λ.

$$P(X=k|\lambda) = \frac{\left(-\frac{\text{Log}\,\eta}{2}\right)^{k} e^{-\left(-\frac{\text{Log}\,\eta}{2}\right)}}{k!}$$

For MLE of η, take derivative wrt η & set to 0.

i.e ~~$\frac{d}{d\eta} = 0$. $\Rightarrow \eta_{MLE} = \frac{d}{d\eta}($~~

Take log likelihood : $L(\eta) = k \operatorname{Log}\left(-\frac{\operatorname{Log} \eta}{2}\right) + \frac{\operatorname{Log} \eta}{2} - \operatorname{Log} k!$

Take derivative and set to zero.

$$\therefore \quad 0 = \frac{d}{d\eta}\left(k \operatorname{Log}\left(-\frac{\operatorname{Log} \eta}{2}\right) + \frac{\operatorname{Log} \eta}{2} - \operatorname{Log} k!\right)$$

$$\frac{k}{\eta \operatorname{Log} \eta} + \frac{1}{2\eta} = 0$$

$$\therefore \boxed{\eta_{MLE} = e^{-2k}.} \quad \therefore \hat{\eta} \text{ is the MLE of } \eta.$$

② Show that: bias of $\hat{\eta}$ is $e^{-2\lambda} - e^{\lambda\left(\frac{1}{e^2} - 1\right)}$

We know that : $\text{bias}(\hat{\eta}) = E[\hat{\eta}] - \eta$. — ①

$$E[\hat{\eta}] = E[e^{-2X}] = \sum_{k=0}^{\infty} e^{-2k} P(X=k) \qquad \{\Sigma k \cdot P(X=k)\}$$

$$P(X=k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \qquad \{\text{poisson's distr}\}$$

$$\therefore \quad E[e^{-2X}] = \sum_{k=0}^{\infty} e^{-2k} \cdot \frac{\lambda^k \cdot e^{-\lambda}}{k!}$$

$$= e^{-\lambda} + \frac{e^{-2}\lambda e^{-\lambda}}{1!} + \frac{e^{-4}\lambda^2 e^{-\lambda}}{2!} + \ldots$$

$$= e^{-\lambda}\left[1 + \frac{e^{-2}\lambda}{1!} + \frac{e^{-4}\lambda^2}{2!} + \ldots\right]$$

$i=1$

Let's put $e^{-2\lambda} = Z$

$\therefore$ ~~BEX~~ $E[e^{-2X}] = e^{-\lambda}\left[1 + \frac{Z}{1!} + \frac{Z^2}{2!} + \cdots\right]$

Now, $1 + Z + \frac{Z^2}{2!} + \cdots$ is Taylor expansion of e^Z.

$\therefore$ $E[e^{-2X}] = \left(\frac{Z^k}{k!}\right) e^{-\lambda} \; = \; e^Z \cdot e^{-\lambda}$.

Repat $Z = e^{-2\lambda}$.

$E[e^{-2X}] = e^{e^{-2}\cdot\lambda} e^{-\lambda}$.

$= e^{\lambda(-1+e^{-2})}$

$= e^{\lambda(\frac{1}{e^2}-1)}$

Subs. in ①.

~~B~~ $\text{bias}(\hat{n}) = E[\hat{n}] - \hat{n}$

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

$\therefore$ $\text{bias}(\hat{n}) = e^{\lambda(\frac{1}{e^2}-1)} - e^{-2\lambda}$.

$$\boxed{\text{bias}(\hat{n}) = -e^{-2\lambda} + e^{\lambda(1/e^2-1)}}.$$

(2.3) ③ we know: $\text{bias}(\hat{n}) = E[\hat{n}] - n$ —— (i)

Here, $\hat{n} = (-1)^X$.

$\therefore$ $E[\hat{n}] = E[(-1)^X] = \sum_{k=0}^{\infty} (-1)^k \cdot \frac{\lambda^k e^{-\lambda}}{k!}$ $\left\{ E(X) = \sum k \cdot P(X=k) \right\}$

$= e^{-\lambda} - \frac{\lambda e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} - \cdots$

$$E[(-1)^x] = e^{-\lambda}\left(1 - \frac{\lambda}{1!} + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \cdots\right)$$

$1 - \lambda + \frac{\lambda^2}{2!} - \cdots$ is Taylor $\exp^n$ of $e^{-\lambda}$.

$$\therefore \; E[(-1)^x] = e^{-\lambda} \cdot e^{-\lambda} = e^{-2\lambda}.$$

Put in ①.

$$\begin{aligned} \text{bias}(\hat{n}) &= E[\hat{n}] - n. \\ &= e^{-2\lambda} - e^{-2\lambda} \\ &= 0. \end{aligned}$$

When bias $= 0$, $\Rightarrow$ estimator is unbiased.

Here, we considered only one observation, which came out to be really close to real value. This means that, though it worked well (unbiased) for 1 sample, doesn't generalize when sample space increases.

(3) Regression and MLE.

(3.1) $y_i = w^T x_i + \epsilon_i$, ϵ_i are $\mathcal{N}(0, \sigma_i^2)$

Taking conditional likelihood.

$$\begin{aligned} P(y \mid x_i, w, \sigma_i) &= \prod_{i=1}^{n} P(y_i \mid x_i, w, \sigma_i) \\ &= \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\,\cancel{2}\sigma_i}\right) \exp\left[\frac{-1}{2\sigma_i^2}(y_i - w^T x_i)^2\right] \end{aligned}$$

Taking log likelihood.

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

2) Calculating MLE.

$$L(w) = \underset{w}{\arg\max} \left(\frac{1}{\sqrt{2\pi}}\right)^n \left[\log\left(\frac{1}{\sigma_i}\right) + \log\left(\exp\left(\sum_1^n \frac{-1}{2\sigma_i^2}(y_i - w^T x_i)^2\right)\right)\right]$$

Take partial derivative wrt. w & set to 0.

~~$\theta = \sum_1^n S_i$~~ First, we can simplify $L(w)$.

consider $s_i = \frac{1}{2\sigma_i^2 \pi}$.

$$\therefore \quad L(w) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left[\log\left(\frac{1}{\sigma_i}\right) + \sum_1^n s_i (y_i - w^T x_i)^2\right]$$

$$0 = \frac{\partial}{\partial w}\left(-\sum_1^n s_i \left(y_i^2 - 2 y_i w^T x_i + w^T . w x_i^2\right)\right)$$

$$0 = \frac{\partial}{\partial w}\left(-s_i \| y - \overset{X^T w}{\cancel{w^T x}} \|^2\right)$$

i.e $\quad -2 X S^T (y - X^T w) = 0$.

$$X S^T y = X S^T X^T w$$

$$\therefore \quad w = (X S^T X^T)^{-1} X S^T y$$

(3.2) $p(\epsilon_i) = \frac{1}{2b} \exp\left(-\frac{|\epsilon_i|}{b}\right)$

$$L(w) = \underset{w}{\arg\min} - \log \prod_1^n P(y_i | x_i; w, b)$$

$$= \underset{w}{\arg\min} - \prod_{i=1}^n \frac{1}{2b} \exp\left(-\frac{|\epsilon_i|}{b}\right)$$

$$= - \prod_{i=1}^n \frac{1}{2b} \exp\left(-\frac{|y_i - w^T x_i|}{b}\right)$$

$$= - \left(\frac{1}{2b}\right)^n \exp\left[\frac{1}{b} \sum_1^n - |y_i - w^T x_i|\right)$$

Q.4.3 Try out your work on synthetic data

1.

Q. Compare Sparsity Pattern of your Lasso solution to the true model of parameters of w*

Sparsity pattern of the Lasso Solution: 57

Sparsity Pattern of the true model w* = 10

Q. Precision: 0.175438596491

Q. Recall: 1

Q. How well are you able to discover the true nonzeros?

The true non-zeros are the first k elements of lasso solution which are non-zero compared with the first k elements of w* which are non-zero. In my case, it is exactly same (k non-zero). So, the ability to discover true non-zeros is 100%

Q. Comment on how Lambda affects these results

For Precision: The value of precision is 1 and then becomes zero as Lambda reaches 0

For Recall: The value of recall increases linearly with lambda until at a point it reaches 1 and then flattens out

2. Q. How are precision and recall affected when sigma=10?

After setting Lambda = 3, the precision and recall values change as follows:

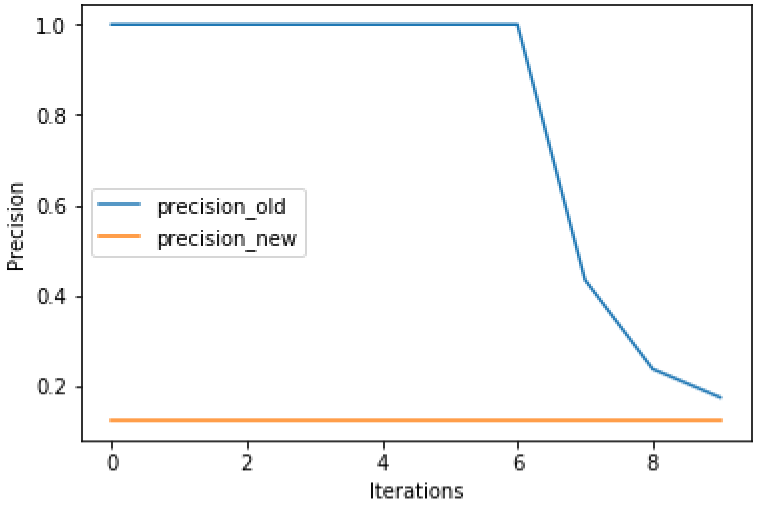
Precision:

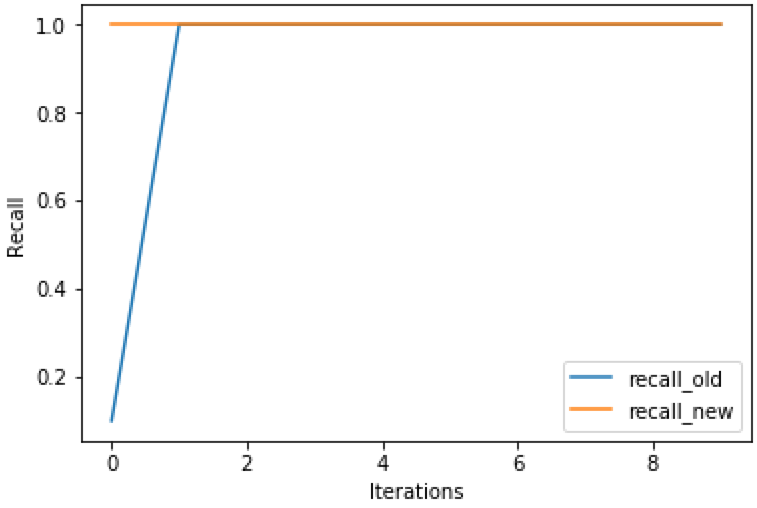
[0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125, 0.125]

Recall:

[1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0]

Once we get an optimal Lambda value, then the precision remains the same (close to) as obtained for sigma = 1. Also, the recall is 1 for all iterations. This means that changing the standard deviation of the noise does not affect the precision or the recall

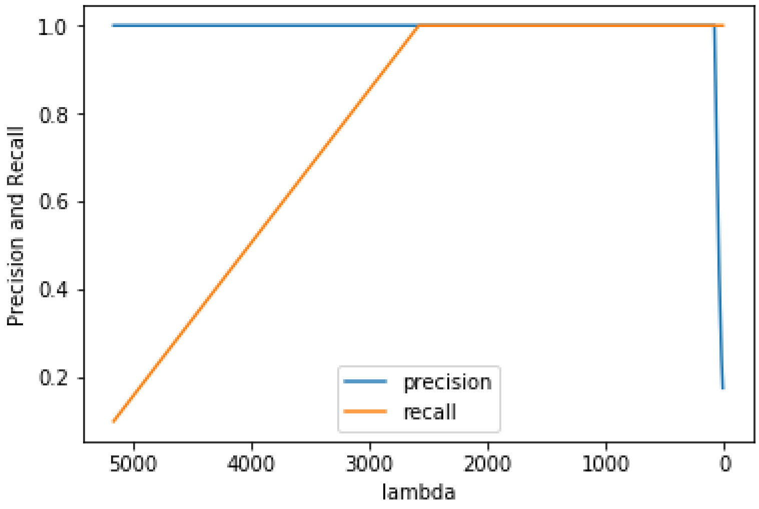




How might you change the Lambda to achieve better precision or recall?

Changing of lambda values does not change the Precision and Recall

Plots:



Q.4.4 Predicting goodness points of a wine given its review

1.

<u>RMSE Training</u>:

[3.6659156152848538, 3.2726658154603703, 2.9817760693142046, 2.7292553433116624, 2.477646256792003, 2.286892526860811, 2.1212885463916873, 1.9524599775424407, 1.8111901053647608, 1.710865888501754, 1.6477442630087837]

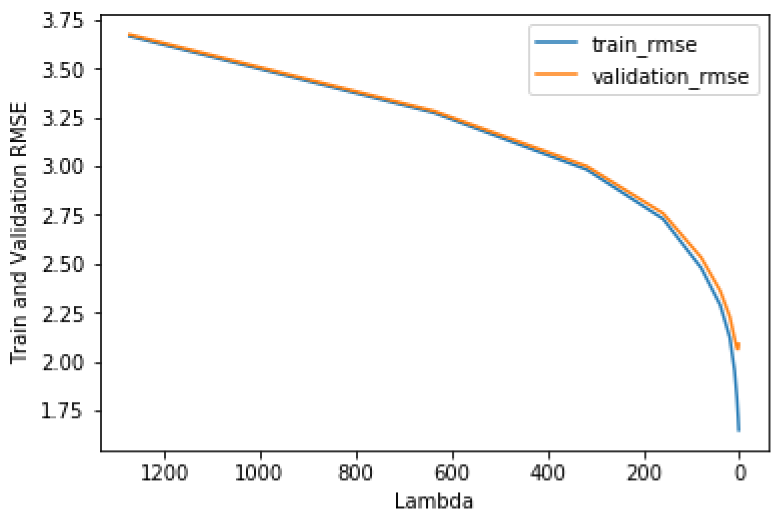
<u>RMSE Validation</u>:

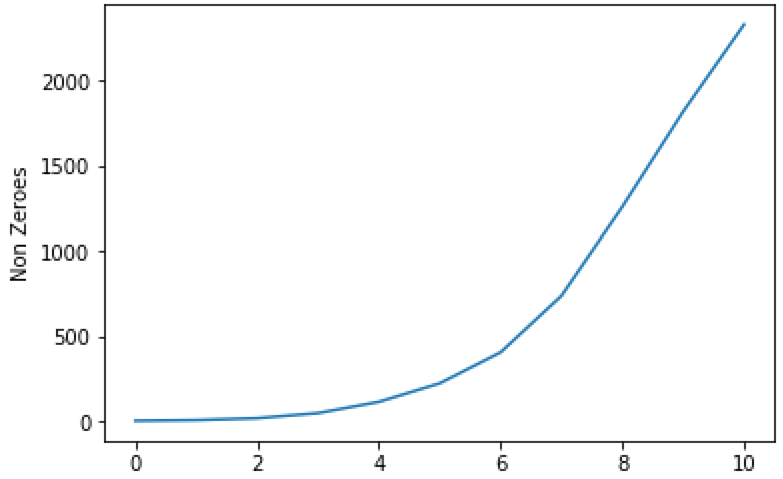
[3.674432595136281, 3.281769756113443, 2.999308942322199, 2.7580247829868267, 2.531297191857193, 2.3604884258078784, 2.2285187942171194, 2.118712422561662, 2.064567203663436, 2.062058813848828, 2.0888203249606496]

<u>Number of Non-Zeros</u>:

[6, 10, 21, 51, 117, 226, 408, 740, 1261, 1820, 2331]

<u>Graphs</u>:





2. Lambda = 2.4841204121093856

   Top 10 Features with highest weight:

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

   Top 10 Features with lowest weight:

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

3. RMSE from Kaggle: 1.94250