Wael Mohammed
271750719

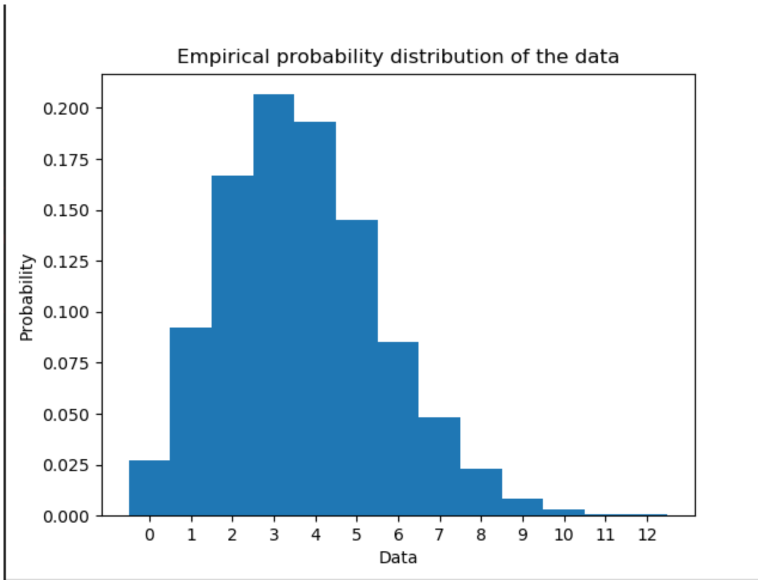
CS 541 – Deep Learning – HW 1

2. Linear Regression via Analytical Solution:-

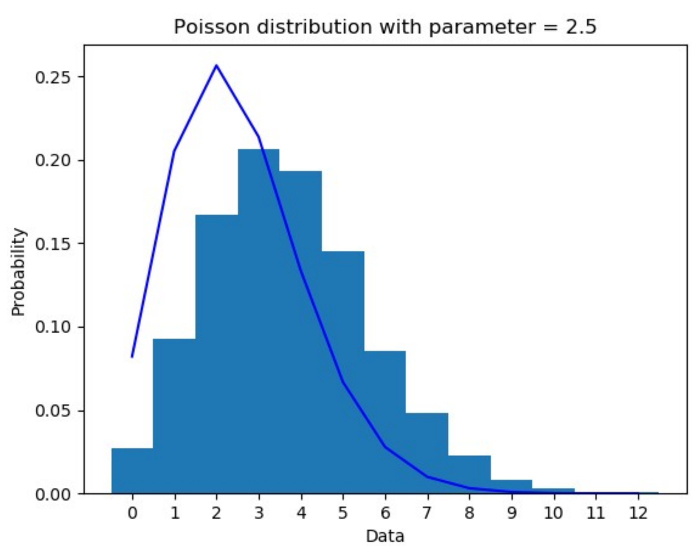
The FMSE on the training data set is 50.47908485415695
The FMSE on the testing data set is 268.7738981451767

3. Probability Distributions:-

a) The histogram portraying the Empirical probability distribution of the data

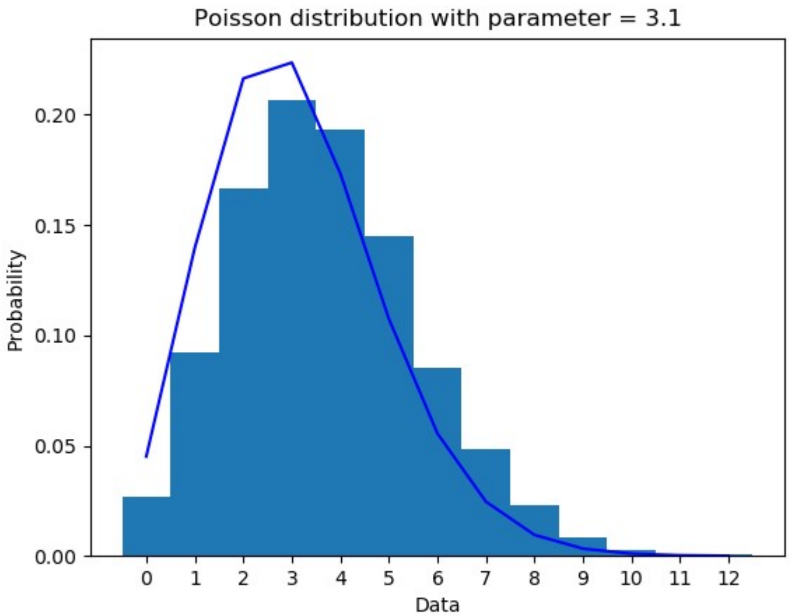


The probability distribution of a Poisson random variable with the rate parameter = 2.5

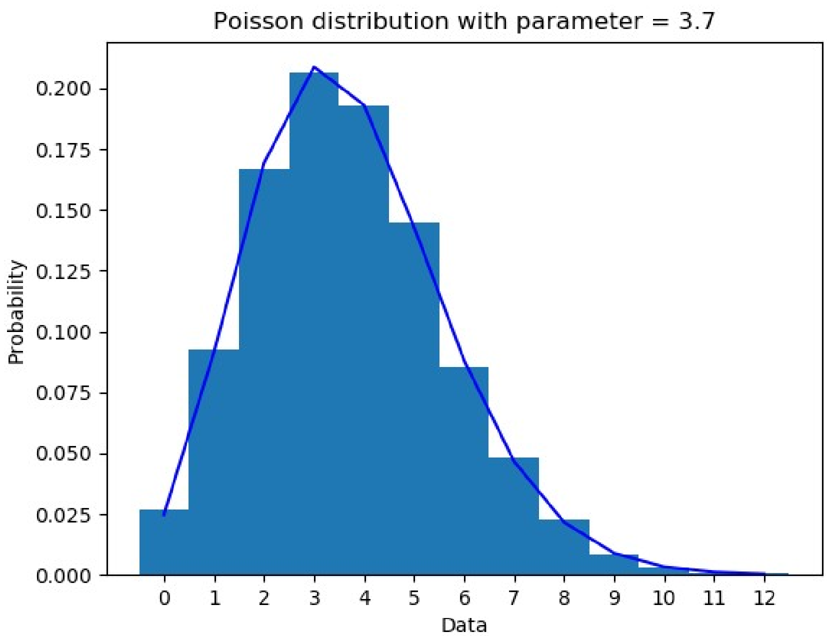


Wael Mohammed
271750719

The probability distribution of a Poisson random variable with the rate parameter = 3.1

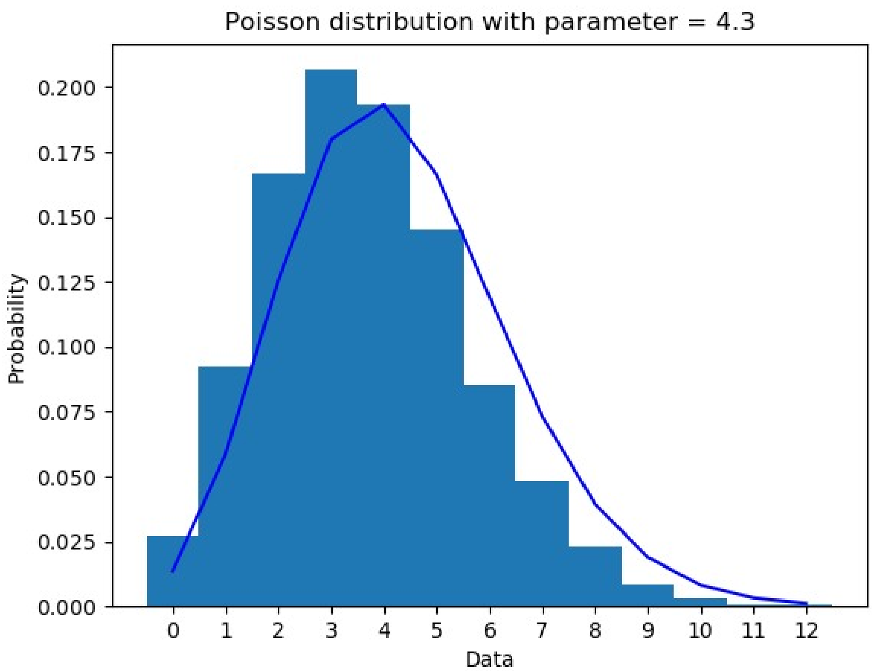


The probability distribution of a Poisson random variable with the rate parameter = 3.7



Wael Mohammed
271750719

The probability distribution of a Poisson random variable with the rate parameter = 4.1



Based on visual inspection the Poisson distribution with parameter = 3.7 is most consistent with the data.

b) i) The value of y tends to be larger with large magnitudes of x (as Mean of y=x^2)
ii) The uncertainty of y tends to be larger with small magnitudes of x (as uncertainty of y = variance of Normal Distribution)

Wael Mohammed
MS RBE
271750719

CS541 - Deep Learning
Homework - I

4. Proofs/Derivations :-

a) Given, $\nabla_x (x^T a) = \nabla_x (a^T x) = a$

Let $x, a \in R^n$ are column vectors,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \text{and} \quad \nabla_x = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

$$x^T a = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} x_1 a_1 + \cdots + x_n a_n \end{bmatrix}$$

$$\text{and } a^T x = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1 x_1 + \cdots + a_n x_n \end{bmatrix}$$

$$\nabla_x (x^T a) = \begin{bmatrix} a_1 + \cdots + 0 \\ \vdots \\ 0 + \cdots + a_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a$$

Similarly, $\nabla_x (a^T x) = \begin{bmatrix} a_1 + \dots + 0 \\ \vdots \\ 0 + \dots + a_n \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a$

$$\therefore \underline{\nabla_x (x^T a) = \nabla_x (a^T x) = a}$$

b) Given, $\nabla_x (x^T A x) = (A + A^T) x$

Let $x \in R^n$ be a column vector and A be $n \times n$ matrix, then,

$x^T = [x_1 \ x_2 \dots x_n]$, $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \dots & & a_{nn} \end{bmatrix}$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$x^T A = [x_1 a_{11} + x_2 a_{21} + \dots + x_n a_{n1} \quad x_1 a_{12} + x_2 a_{22} + \dots + x_n a_{n2} \dots$
$\dots x_1 a_{1n} + \dots + x_n a_{nn}]$

$x^T A x = [x_1 a_{11} + \dots + x_n a_{n1} \quad x_1 a_{12} + \dots + x_n a_{n2} \dots x_1 a_{1n} + \dots + x_n a_{nn}] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$x^T A x = [(x_1^2 a_{11} + x_1 x_2 a_{21} + \dots + x_1 x_n a_{n1}) + (x_2 x_1 a_{12} + x_2^2 a_{22} + \dots + x_2 x_n a_{n2}) +$
$\dots + (x_n x_1 a_{1n} + x_n x_2 a_{2n} + \dots + x_n^2 a_{nn})]$

Scanned with CamScanner

$$\nabla_x x^T A x = \begin{bmatrix} \frac{\partial (x^T A x)}{\partial x_1} \\ \frac{\partial (x^T A x)}{\partial x_2} \\ \vdots \\ \frac{\partial (x^T A x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} (2x_1 a_{11} + x_2 a_{21} + \dots + x_n a_{n1}) + (x_2 a_{12}) + \dots \\ \dots + x_n a_{1n} \\ (x_1 a_{21}) + (x_1 a_{12} + 2x_2 a_{22} + \dots + x_n a_{n2}) + \\ \dots + x_n a_{2n} \\ \vdots \\ (x_1 a_{n1}) + (x_2 a_{n2}) + \dots + \\ (x_1 a_{1n} + x_2 a_{2n} + \dots + 2x_n a_{nn}) \end{bmatrix}$$

On re-arranging the terms by taking commons, we get,

$$\nabla_x x^T A x = \begin{bmatrix} 2x_1 a_{11} + x_2(a_{12} + a_{21}) + \dots + x_n(a_{1n} + a_{n1}) \\ x_1(a_{21} + a_{12}) + 2x_2 a_{22} + \dots + x_n(a_{2n} + a_{n2}) \\ \vdots \\ x_1(a_{n1} + a_{1n}) + x_2(a_{n2} + a_{2n}) + \dots + 2x_n a_{nn} \end{bmatrix} \quad \hookrightarrow (1)$$

$= L.H.S$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \cdots & & a_{nn} \end{bmatrix} \qquad A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & & \vdots \\ \vdots & & \ddots & \\ a_{1n} & \cdots & & a_{nn} \end{bmatrix}$$

$$A+A^T = \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{n1} \\ a_{21}+a_{12} & 2a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1}+a_{1n} & - & - & 2a_{nn} \end{bmatrix}$$

$$(A+A^T)x = \begin{bmatrix} 2a_{11} & a_{12}+a_{21} & \cdots & a_{1n}+a_{n1} \\ a_{21}+a_{12} & 2a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1}+a_{1n} & - & - & 2a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\cancel{2a_{11}x}$$

$$(A+A^T)x = \begin{bmatrix} 2x_1a_{11} + x_2(a_{12}+a_{21}) + \cdots + x_n(a_{1n}+a_{n1}) \\ x_1(a_{21}+a_{12}) + 2x_2a_{22} + \cdots + x_n(a_{2n}+a_{n2}) \\ \vdots \\ x_1(a_{n1}+a_{1n}) + x_2(a_{n2}+a_{2n}) + \cdots + 2x_na_{nn} \end{bmatrix}$$

↳ (2)

$= R.H.S$

On comparing (1) and (2), we get L.H.S = R.H.S,

$$\therefore \nabla_x(x^TAx) = (A+A^T)x$$

i) Given, $\nabla_x (x^T A x) = 2Ax$

Let $x \in \mathbb{R}^n$ be a column vector and A be a symmetric $n \times n$ matrix, then,

$$L.H.S = \begin{bmatrix} 2x_1 a_{11} + x_2(a_{12}+a_{21}) + \dots + x_n(a_{1n}+a_{n1}) \\ x_1(a_{21}+a_{12}) + 2x_2 a_{22} + \dots + x_n(a_{2n}+a_{n2}) \\ \vdots \\ x_1(a_{n1}+a_{1n}) + x_2(a_{n2}+a_{2n}) + \dots + 2x_n a_{nn} \end{bmatrix}$$

(From 1)

We know that in a symmetric matrix $a_{ij} = a_{ji}$ for $i \neq j$, then,

$$L.H.S = \begin{bmatrix} 2x_1 a_{11} + 2x_2 a_{12} + \dots + 2x_n a_{1n} \\ 2x_1 a_{21} + 2x_2 a_{22} + \dots + 2x_n a_{2n} \\ \vdots \\ 2x_1 a_{n1} + 2x_2 a_{n2} + \dots + 2x_n a_{nn} \end{bmatrix}$$

$$= 2Ax = R.H.S$$

Thus the L.H.S = R.H.S,

$$\therefore \nabla_x (x^T A x) = 2Ax$$

d) Given, $\nabla_x \left[(Ax+b)^T(Ax+b)\right] = 2A^T(Ax+b)$

Let $x \in \mathbb{R}^n$ is a column vector, A is $n \times n$ symmetric matrix and $b \in \mathbb{R}^n$ is a constant column vector, then,

$$(Ax+b)^T = (Ax)^T + b^T \qquad (\because (x+y)^T = x^T + y^T)$$
$$= x^T A^T + b^T \qquad (\because (xy)^T = y^T x^T)$$

$$(Ax+b)^T(Ax+b) = (x^T A^T + b^T)(Ax+b)$$
$$= x^T A^T A x + x^T A^T b + b^T A x + b^T b$$

Let $\underset{n\times n}{M} = A^T A$ and Let $\underset{1\times n}{N^T} = b^T A$ and $\underset{n\times 1}{N} = A^T b$

$$(Ax+b)^T(Ax+b) = x^T M x + x^T A^T b + \cancel{b^T A x}\, N^T x + b^T b$$

$$\nabla_x \left[(Ax+b)^T(Ax+b)\right] = \nabla_x x^T M x + \nabla_x \underset{x^T N}{\cancel{x^T A^T b}} + \nabla_x N^T x + \nabla_x b^T b$$

$$= 2Mx + N + N + 0 \quad (\text{from 4.a) and 4.c)})$$
$$= 2Mx + 2N$$
$$= 2A^T A x + 2A^T b$$
$$= 2A^T(Ax+b)$$
$$= \underline{\underline{R.H.S}}$$

$\therefore \underline{\underline{\nabla_x \left[(Ax+b)^T(Ax+b)\right] = 2A^T(Ax+b)}}$ ($\because$ L.H.S = R.H.S)