Name: Samarth Shah (590146168), Fenil Desai (249812226)
Date: 01/19/2022
Subject: CS 541 Deep Learning

Deep Learning HW #1

Contents

Problem 1: Python and Numpy Warm-up Exercises 2

Problem 2: Linear Regression via Analytical Solution 2

a. Age Regressor: 2

Problem 3: Probability Distributions 3

a. Estimating the Parameters of a Probability Distribution: 3

b. Conditional Probability Distributions to Represent the Uncertainty of Functions: 4

Problem 4: Proofs and Derivation 4

a. Prove $\nabla_x(x^T a) = \nabla_x(a^T x) = a$ 4

b. Prove $\nabla_x(x^T Ax) = (A + A^T)x$ 7

c. Prove $\nabla_x(x^T Ax) = 2Ax$ 9

d. Prove $\nabla_x((Ax + b)^T * (Ax + b)) = 2A^T(Ax + b)$ 9

Name: Samarth Shah (590146168), Fenil Desai (249812226)
Date: 01/19/2022
Subject: CS 541 Deep Learning

Problem 1: Python and NumPy Warm-up Exercises

Please refer to ***homework1_template.py*** for solutions of 1(a) to 1(n).

Problem 2: Linear Regression via Analytical Solution

a. Age Regressor:

- We have implemented linear regression via analytical solution using linear algebraic operations in NumPy. Find the methods *linear_regression* and *train_age_regressor* in ***homework1_template.py.***
    - Note: Before running the script, please include the datasets in the same folder.
- Analytical Solution to obtain the weights (w):

$$f_{MSE}(y; \hat{y}; w) = \frac{1}{2n} \sum_{i=1}^{n} \left(g(x^{(i)}; w) - y^{(i)} \right)^2$$

$$= \frac{1}{2n} \sum_{i=1}^{n} \left(x^{(i)T} w - y^{(i)} \right)^2$$

→ Solving for w

$$\nabla_w f_{MSE}(y; \hat{y}; w) = \nabla_w \left[\frac{1}{2n} \sum_{i=1}^{n} \left(x^{(i)T} w - y^{(i)} \right)^2 \right]$$

$$= \frac{1}{2n} \sum_{i=1}^{n} \nabla_w \left[\left(x^{(i)T} w - y^{(i)} \right)^2 \right]$$

$$= \frac{1}{n} \sum_{i=1}^{n} x^{(i)} \left(x^{(i)T} w - y^{(i)} \right)$$

→ By setting to 0, splitting the sum apart, & solving, we reach the solution:

$$\nabla_w f_{MSE}(y; \hat{y}; w) = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} \left(x^{(i)T} w - y^{(i)} \right)$$

$$0 = \sum_{i} x^{(i)} x^{(i)T} w - \sum_{i} x^{(i)} y^{(i)}$$

$$\sum_i x^{(i)} x^{(i)T} w = \sum_i x^{(i)} y^{(i)}$$

$$w = \left(\sum_i x^{(i)} x^{(i)T}\right)^{-1} \sum_i x^{(i)} y^{(i)}$$

$$\therefore \boxed{w = (XX^T)^{-1} X y} \quad \text{(where, X is a design matrix)}$$

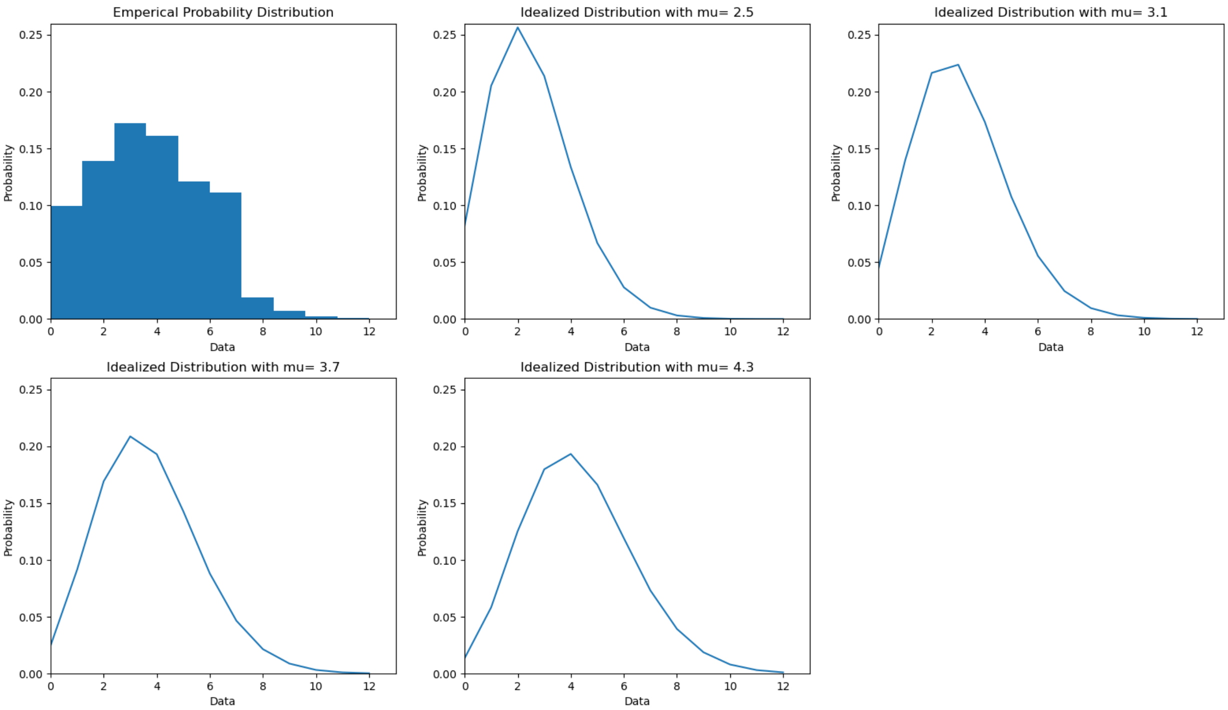
- After optimizing the weights on training set, we then computed the cost f_{MSE} on both testing and training dataset. The output is reported below:

```
Training Loss: 50.46582283534778
Testing Loss: 268.7869643405511
```

Problem 3: Probability Distributions

a. **Estimating the Parameters of a Probability Distribution:**

- We have plotted the empirical probability of the data in ***PoissonX.npy.***
- Next, we plotted the probability distribution of a Poisson random variable with the following rate parameters: 2.5, 3.1, 3.7, and 4.3.
- The plots are as shown below:

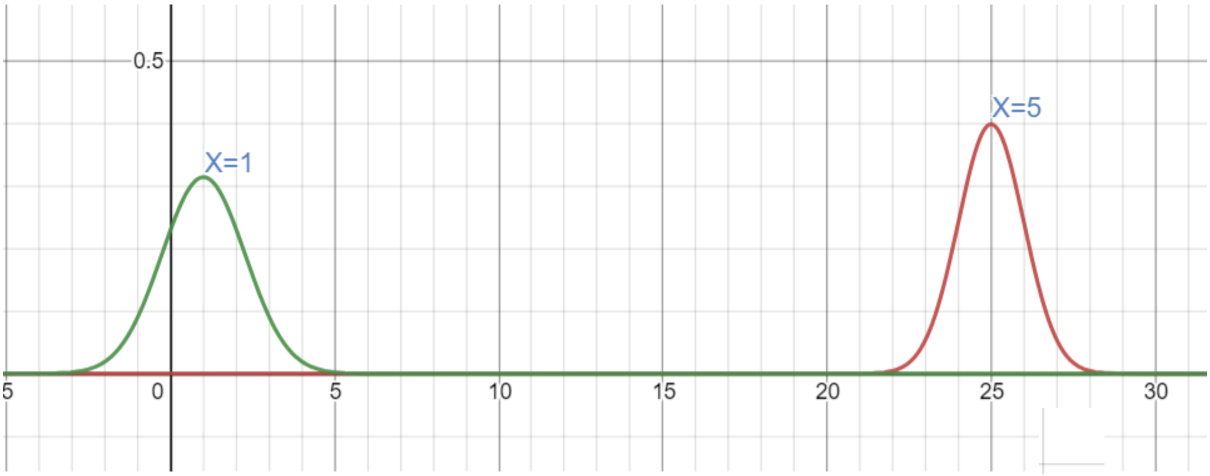


- Based on observations the parameter value of mu=3.7 is most consistent with the data.

b. Conditional Probability Distributions to Represent the Uncertainty of Functions:

i. The corresponding value of *y* tend to be larger for the values of *x* with large magnitude as mean of the given normal distribution varies according to x^2. We can observe in the graph below that at x=5 the mean of distribution is 25 therefore indicating that y would be large.

ii. The uncertainty in the corresponding value of y tend to be larger for the values of x with small magnitude. We can observe in the graph below that at x=1 the spread is more when compared to x=5. More spread indicates that the uncertainty is higher for lower value of x.



Problem 4: Proofs and Derivation

a. Prove $\nabla_x(x^T a) = \nabla_x(a^T x) = a$

$\rightarrow$ let x be a column vector $x = [x_1, x_2, --- x_n]^T$

let a be a column vector $a = [a_1, a_2 ---- a_n]^T$

now;

$$x^T a = [x_1, x_2, x_3 --- x_n] \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}$$

$$\Rightarrow x^T a = (a_1 x_1 + a_2 x_2 + a_3 x_3 ---- a_n x_n) \quad \text{—①}$$

Similarly;

$$a^T x = [a_1, a_2, a_3 \cdots a_n] \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$\Rightarrow a^T x = (a_1 x_1 + a_2 x_2 + a_3 x_3 \cdots a_n x_n) \quad — (2)$$

now, we know that;

$$\nabla_x f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

→ now, let us calculate, $\frac{\partial (x^T a)}{\partial x_1}$

$$\frac{\partial (x^T a)}{\partial x_1} = a_1 + 0 + 0 \cdots 0$$

Similarly

$$\frac{\partial (x^T a)}{\partial x_2} = 0 + a_2 + \cdots 0 \quad \& \quad \frac{\partial (x^T a)}{\partial x_n} = a_n$$

$$\therefore \boxed{\nabla_x (x^T a) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a}$$

Similarly;

$$\frac{\partial(a^T x)}{\partial x_1} = a_1 + 0 + 0 - - - 0$$

$$\frac{\partial(a^T x)}{\partial x_2} = 0 + a_2 + 0 - - - 0$$

$$\frac{\partial(a^T x)}{\partial x_n} = a_n$$

$$\therefore \boxed{\nabla_x (a^T x) = \begin{bmatrix} a_1 \\ a_2 \\ | \\ a_n \end{bmatrix} = a}$$

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

b. Prove $\nabla_x(x^T A x) = (A + A^T)x$

→ Prove that: $\nabla_x(x^T A x) = (A + A^T)x$

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

let us represent A as → $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ or $[b_1 \ b_2 \ \cdots \ b_n]$

where, $a_1 \cdots a_n$ are the rows of A.
where, $b_1 \cdots b_n$ are the columns of A.
So, we can represent Ax as $\begin{bmatrix} a_1 x \\ a_2 x \\ \vdots \\ a_n x \end{bmatrix}$

→ now;

$$x^T A x = x^T [a_1 x \ \ a_2 x \ \cdots \ a_n x]^T$$

$$= x_1 a_1 x + x_2 a_2 x \cdots x_n a_n x$$

where, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

→ now, let us take derivative for one component x_k:

$$\frac{\partial (x^T A x)}{\partial x_k}$$

$$\frac{\partial(x^T A x)}{\partial x_k} = a_k x + a_{1k} x_1 + a_{2k} x_2 + \cdots \cdots a_{nk} x_n$$

$$= a_k x + \sum_{i=1}^{n} a_{ik} x_i$$

(rows × x) (colums × x)

$$= a_k x + (b_k)^T x$$

$$\therefore \nabla_x (x^T A x) = \begin{bmatrix} a_1 x + (b_1)^T x \\ a_2 x + (b_2)^T x \\ \vdots \\ a_n x + (b_n)^T x \end{bmatrix}$$

$$\therefore \begin{bmatrix} a_1 x \\ a_2 x \\ \vdots \\ a_n x \end{bmatrix} + \begin{bmatrix} (b_1)^T x \\ (b_2)^T x \\ \vdots \\ (b_n)^T x \end{bmatrix}$$

$$\therefore \boxed{\begin{aligned} \nabla_x (x^T A x) &= Ax + A^T x \\ &= (A + A^T) x \end{aligned}}$$

c. Prove $\nabla_x(x^T A x) = 2Ax$

Given A is symmetric $n \times n$ matrix

$\therefore A = A^T$

$\therefore$ from (b) we know $\nabla_x (x^T A x) = (A + A^T)x$

$$\boxed{\therefore A = A^T}$$

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

Hence Proved //

d. Prove $\nabla_x((Ax + b)^T * (Ax + b)) = 2A^T(Ax + b)$

LHS :- $\nabla_x \left[(x^T A^T + b^T)(Ax + b) \right]$

$$\Rightarrow \nabla_x \left[x^T (A^T A) x + x^T (A^T b) + (b^T A) x + b^T b \right]$$

Let us consider $A^T A = K$ (n×n matrix)

$A^T b = a$ → vector (n×1)

$b^T A = a^T$

$$\Rightarrow \nabla_x \left[x^T K x + x^T a + a^T x + b^T b \right]$$

$\therefore \nabla_x b^T b = 0$, $\nabla_x (x^T A x) = 2Ax$, $\nabla_x (x^T a) = \nabla_x (a^T x) = a$

$$\Rightarrow \nabla_x \quad 2Kx + a + a + 0$$

$$\Rightarrow 2A^T A x + 2A^T b \Rightarrow 2A^T (Ax + b) = \text{RHS} //$$

End Of Assignment