Name:

University of Illinois Spring 2020

CS 446/ECE 449 Machine Learning
Homework 1: Linear Regression

Due on Thursday February 6 2020, noon Central Time

1. **[17 points]** Linear Regression

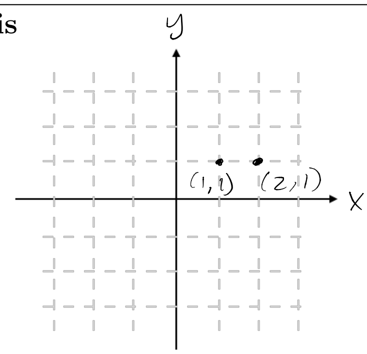
We are given a dataset $\mathcal{D} = \{(1,1), (2,1)\}$ containing two pairs (x, y), where each $x \in \mathbb{R}, y \in \mathbb{R}$ denotes a real-valued number.

We want to find the parameters $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in \mathbb{R}^2$ of a linear regression model $\hat{y} = w_1 x + w_2$ using

$$\min_w \frac{1}{2} \sum_{(x,y)\in\mathcal{D}} \left(y - w^\top \begin{bmatrix} x \\ 1 \end{bmatrix} \right)^2. \quad (1)$$

(a) (2 points) Plot the given dataset and find the optimal w^* by inspection.

Your answer: **Mark the axis**



$w^* = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

(b) (4 points) Using general matrix vector notation, the program in Eq. (1) is equivalent to

$$\min_w \frac{1}{2} \|\mathbf{y} - \mathbf{X}w\|_2^2. \quad (2)$$

Specify the dimensions of the introduced matrix $\mathbf{X}$ and the introduced vector $\mathbf{y}$. Also write down explicitly the matrices and vectors using the values in the given dataset $\mathcal{D}$.

Your answer: m = # of samples n: # features.

$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ x_{m1} & \cdots & x_{mn} & 1 \end{bmatrix}$

col of 1's for bias

$X \in \mathbb{R}^{m \times (n+1)}$

$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$

$y \in \mathbb{R}^m$

This data set:

$X = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$ $x \in \mathbb{R}^{2\times 2}$

$y = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $y \in \mathbb{R}^2$

Name: ______________________

(c) (3 points) **Derive** the general analytical solution for the program given in Eq. (2). Also plug in the values for the given dataset $\mathcal{D}$ and compute the solution numerically.

Your answer:

$\frac{1}{2}\|Y - X^T w\|_2^2$

$\frac{1}{2}(Y - X^T w)^T (Y - X^T w)$

$\frac{1}{2}(Y^T - w^T X)(Y - X^T w)$

$\frac{1}{2}(\underbrace{Y^T Y - Y^T X^T w - w^T X Y + w^T X X^T w}_{})$

1x1 matricies so $Y^T X^T w = w^T X Y$

$\frac{1}{2}(Y^T Y - 2 w^T X Y + w^T X X^T w$

$\frac{1}{2} Y^T Y - w^T X Y + \frac{1}{2} w^T X X^T w$

↓ differentiate w/ respect to w

$-XY + XX^T w^* = 0$

$XX^T w^* = XY$

$\boxed{w^* = (XX^T)^{-1}(XY)}$

Note: here X^T is the X defined in # 2.

$(XX^T) = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$

$(XX^T)^{-1} = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$

$XY = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$w^* = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

$\boxed{= \begin{bmatrix} 0 \\ 1 \end{bmatrix}}$

(d) (1 point) Numerous ways exist to compute this solution via PyTorch. Read the docs for the functions ‘torch.gels’, ‘torch.gesv’, and ‘torch.inverse’. Use all three approaches when completing the file A1_LinearRegression.py and verify your answer. Which solution provides the most accurate value for w_1 for our dataset?

Your answer:

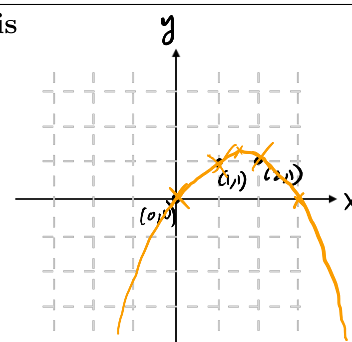
torch.solve (this replaced torch.gesv).

(e) (6 points) We are now given a dataset $\mathcal{D} = \{(0,0),(1,1),(2,1)\}$ of pairs (x, y) with $x, y \in \mathbb{R}$ for which we want to fit a quadratic model $\hat{y} = w_1 x^2 + w_2 x + w_3$ using the program given in Eq. (2). Specify the dimensions of the matrix $\mathbf{X}$ and the vector $\mathbf{y}$. Also write down explicitly the matrix and vector using the values in the given dataset. Find the optimal solution w^* and draw it together with the dataset into a plot.

Your answer: **Mark the axis**

$X = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \quad \mathbb{R}^{3\times 3}$

$y = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$



black = what our dataset is
orange = what our model predicts.

$XY = \begin{bmatrix} 0 & 1 & 4 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$

$= \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix}$

$y = -\frac{1}{2}x^2 + \frac{3}{2}x$

$w^* = \begin{bmatrix} 3/2 & -3 & \frac{1}{2} \\ -3 & 13/2 & -3/2 \\ \frac{1}{2} & -3/2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix}$

$w^* = (XX^T)^{-1}(XY)$

$(XX^T)^{-1} = \left(\begin{bmatrix} 0 & 1 & 4 \\ 0 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \right)^{-1} = \left(\begin{bmatrix} 17 & 9 & 5 \\ 9 & 5 & 3 \\ 5 & 3 & 3 \end{bmatrix} \right)^{-1} = \begin{bmatrix} 3/2 & -3 & \frac{1}{2} \\ -3 & 13/2 & -3/2 \\ \frac{1}{2} & -3/2 & 1 \end{bmatrix}$

$\boxed{w^* = \begin{bmatrix} -1/2 \\ 3/2 \\ 0 \end{bmatrix}}$

(f) (1 points) Complete A2_LinearRegression2.py and verify your reply for the previous answer. How did you specify the matrix $\mathbf{X}$?

Your answer:

```
X = torch.Tensor([[0,0,1],[1,1,1],[4,2,1]])   res1 = torch.solve(y, X)
y = torch.Tensor([[0],[1],[1]])               w* = res1[0].
```