In the name of God

HW2

Deadline: NOV 17th

Email: mlfall2023@gmail.com

About submission:

- File's naming format:
  - For Homeworks: "HWX _Student NO."
    Exe: HW2 _400613198.pdf
  - For MATLAB Homeworks: "PRJX _Student NO."
    Exe: PRJ1 _400613198.pdf
- About late submissions: %10 of daily reduction of the HW mark.

Exercise 7.2 Multi-output linear regression

(Source: Jaakkola.)

When we have multiple independent outputs in linear regression, the model becomes

$$p(\mathbf{y}|\mathbf{x}, \mathbf{W}) = \prod_{j=1}^{M} \mathcal{N}(y_j|\mathbf{w}_j^T \mathbf{x}_i, \sigma_j^2)$$

Since the likelihood factorizes across dimensions, so does the MLE. Thus

$$\hat{\mathbf{W}} = [\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_M]$$

where $\hat{\mathbf{w}}_j = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{Y}_{:,j}$.

In this exercise we apply this result to a model with 2 dimensional response vector $\mathbf{y}_i \in \mathbb{R}^2$. Suppose we have some binary input data, $x_i \in \{0, 1\}$. The training data is as follows:

x	y
0	$(-1, -1)^T$
0	$(-1, -2)^T$
0	$(-2, -1)^T$
1	$(1, 1)^T$
1	$(1, 2)^T$
1	$(2, 1)^T$

Let us embed each x_i into 2d using the following basis function:

$$\boldsymbol{\phi}(0) = (1, 0)^T, \quad \boldsymbol{\phi}(1) = (0, 1)^T$$

The model becomes

$$\hat{\mathbf{y}} = \mathbf{W}^T \boldsymbol{\phi}(x)$$

where $\mathbf{W}$ is a 2×2 matrix. Compute the MLE for $\mathbf{W}$ from the above data.

Exercise 7.3 Centering and ridge regression

Assume that $\overline{\mathbf{x}} = 0$, so the input data has been centered. Show that the optimizer of

$$J(\mathbf{w}, w_0) = (\mathbf{y} - \mathbf{X}\mathbf{w} - w_0\mathbf{1})^T(\mathbf{y} - \mathbf{X}\mathbf{w} - w_0\mathbf{1}) + \lambda\mathbf{w}^T\mathbf{w}$$

is

$$\begin{aligned} \hat{w}_0 &= \overline{y} \\ \mathbf{w} &= (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y} \end{aligned}$$

Exercise 7.5 MLE for the offset term in linear regression

Linear regression has the form $\mathbb{E}[y|\mathbf{x}] = w_0 + \mathbf{w}^T\mathbf{x}$. It is common to include a column of 1's in the design matrix, so we can solve for the offset term w_0 term and the other parameters $\mathbf{w}$ at the same time using the normal equations. However, it is also possible to solve for $\mathbf{w}$ and w_0 separately. Show that

$$\hat{w}_0 = \frac{1}{N}\sum_i y_i - \frac{1}{N}\sum_i \mathbf{x}_i^T\mathbf{w} = \overline{y} - \overline{\mathbf{x}}^T\mathbf{w}$$

So $\hat{w}_0$ models the difference in the average output from the average predicted output. Also, show that

$$\hat{\mathbf{w}} = (\mathbf{X}_c^T\mathbf{X}_c)^{-1}\mathbf{X}_c^T\mathbf{y}_c = \left[\sum_{i=1}^{N}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^T\right]^{-1}\left[\sum_{i=1}^{N}(y_i - \overline{y})(\mathbf{x}_i - \overline{\mathbf{x}})\right]$$

where $\mathbf{X}_c$ is the centered input matrix containing $\mathbf{x}_i^c = \mathbf{x}_i - \overline{\mathbf{x}}$ along its rows, and $\mathbf{y}_c = \mathbf{y} - \overline{\mathbf{y}}$ is the centered output vector. Thus we can first compute $\hat{\mathbf{w}}$ on centered data, and then estimate w_0 using $\overline{y} - \overline{\mathbf{x}}^T\hat{\mathbf{w}}$.

Q4) Prove that each of these distributions belong to the exponential family

a) Normal
b) Poisson
c) Gamma
d) Binomial with certain number f trials
e) Chi-squared

Good luck