Keertana V. Chidambaram
12211266

Pset 1, TTIC 31020.

1. Linear Regression.

Problem 1: Short ans: YES Explanation:

Assumption: $(X^T X)$ is invertible.

We know,

$$w^* = (X^T X)^{-1} X^T y.$$

$$w' = (X^T X)^{-1} X^T [a\vec{y} + \vec{b}]$$

where $a \in \mathbb{R}$ $\vec{b} = \underset{n \times 1}{\begin{bmatrix} b \\ b \\ \vdots \\ b \end{bmatrix}} = b \times \underset{n \times 1}{\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}}$

$$w' = a[(X^T X)^{-1} X^T y] + (X^T X)^{-1} X^T . b$$

$$= a w^* + (X^T X)^{-1} X^T b$$

Let $u = (X^T X)^{-1} X^T b.$

$\Rightarrow (X^T X) u = X^T b$

$\Rightarrow X X^T X u = X X^T b$

$\Rightarrow (X X^T)^{-1} (X X^T) X u = b.$

$\Rightarrow Xu = b$.

on expanding, let $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_d \end{bmatrix}_{d \times 1}$

$$X = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1d} \\ \vdots & \vdots & & \\ 1 & x_{n2} & \cdots & x_{nd} \end{bmatrix}_{n \times d}$$

first columns is ones for the intercept

So,

$$u_1 + x_{i2} u_2 + \cdots + x_{id} u_d = 0. \quad \forall i \in \{1, 2, \ldots, n\}.$$

Since u_is are fixed and this equation holds for all n observations, we can infer that $u_1 = b$ and $u_i = 0 \ \forall i \neq b$.

Thus $u = \begin{bmatrix} b \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$w' = aw^* + u$.

So w' is defined s.T $w'_1 = aw_1^* + b$ (intercept term)

and $w'_i = aw'_i \quad \forall i \in \{2, 3, \ldots, d\}$.

Problem 2 : Yes.

$\cancel{X = y, w^*}$ $\quad Xw^* = y$ and

~~and~~ $\cancel{X' = y\tilde{w}}$ $\quad X'\tilde{w} = y$.

So, $\quad Xw^* = X'\tilde{w}$.

$X' = XC$, where C is a diagonal matrix such that

$$C = \begin{bmatrix} c_1 & 0 & 0 & \cdots & 0 \\ 0 & c_2 & 0 & \cdots & 0 \\ 0 & & \ddots & & \vdots \\ \vdots & & & \ddots & \\ 0 & 0 & \cdots & & c_d \end{bmatrix}$$

$d \times d$.

So. $\quad Xw^* = XC\tilde{w}$

$$\Rightarrow \tilde{w} = C^{-1}w^*$$

$$C^{-1} = \begin{bmatrix} 1/c_1 & 0 & \cdots & 0 \\ 0 & 1/c_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/c_d \end{bmatrix} \quad \text{so, } \tilde{w} = \begin{bmatrix} w_1^*/c_1 \\ w_2^*/c_2 \\ \vdots \\ w_d^*/c_d \end{bmatrix}$$

Problem 3: No.

Maximizing the log likelihood is the same as maximizing:

$$\cancel{\max W^* = \underset{w}{\arg\max} \frac{1}{n}}$$

$$w^* = \underset{w}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \left[-\frac{(y_i - f(x_i, w))^2}{2\sigma_{x_i}^2} - \log \sigma_{x_i}\sqrt{2\pi} \right]$$

$$= \underset{w}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{-(y_i - w \cdot x_i)^2}{2\sigma_{x_i}^2} - \log \sigma_{x_i}\sqrt{2\pi} \right].$$

$\sigma_{x_i}^2$ cannot be taken out of the summation, hence we cannot calculate $W^*_{\cancel{max}}$ without knowing σ_{x_i} because W^* depends on it.

Problem 4: Yes.

$$w^* = \underset{w}{\operatorname{argmax}} \ \frac{1}{n} \sum_{i=1}^{n} \left[-\frac{(y_i - wx_i)^2}{2\sigma_{x_i}^2} - \log \sigma_{x_i} \sqrt{2\pi} \right]$$

only this part of the function is dependent on w so

$$w^* = \operatorname{argmax} \sum_{i=1}^{n} -\frac{(y_i - wx_i)^2}{2\sigma_{x_i}^2}.$$

$$= \operatorname{argmax} \sum_{i=1}^{n} -\left(\frac{y_i}{\sigma_{x_i}\sqrt{2}} - \frac{wx_i}{\sigma_{x_i}\sqrt{2}} \right)^2.$$

since we know σ_{x_i}, w^* is the optimal parameters we get when we OLS regress $\frac{y_i}{\sigma_{x_i}\sqrt{2}}$ and $\frac{x_i}{\sigma_{x_i}\sqrt{2}}$.