CREED REILLY 6721694310

EE559 HW04

06/28/2021

Check... may need work

1) $\Gamma(2,1)$ PDF: $p(z) = z\exp(-z)$, $z > 0$

Poisson pmf: $p_x(x) = \lambda^x \exp(-\lambda)/x!$, $\lambda > 0$, $x = 0, 1, \ldots$

$p_{x|z}(x|z)$ iid $\{x_1, \ldots, x_n\}$

MAP: $\underset{\lambda}{\operatorname{argmax}}\ p(w|x) = \underset{\lambda}{\operatorname{argmax}}\ \dfrac{p(x|z)\,p(z)}{p(x)}$

$$= \underset{\lambda}{\operatorname{argmax}}\ p(\underline{x}|z)\,p(z)$$

$$= \underset{\lambda}{\operatorname{argmax}} \prod_{i=1}^{n} \left(\lambda^{x_i} \exp(-\lambda)/x_i!\right) z\exp(-z)$$

$$= \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \left[x_i \ln(\lambda) - \lambda - \ln(x_i!)\right] + \ln(z) - z$$

$$= \underset{\lambda}{\operatorname{argmax}} \left[\ln(\lambda)\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \ln(x_i!) + \ln(z) - z\right] \quad \text{— (1)}$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial (1)}{\partial \lambda} = \frac{n\mu}{\lambda} - n = 0 \rightarrow \frac{\mu}{\lambda} = 1 \rightarrow \boxed{\lambda = \mu} > 0$$

$$\frac{\partial^2 (1)}{\partial \lambda^2} = \frac{-n\mu}{\lambda^2} = \frac{-n\mu}{\mu^2} = \boxed{\frac{-n}{\mu} < 0} \leftarrow \text{maximum!}$$

QED

2) $P_{X|\lambda}(x|\lambda) = \lambda^x \exp(-\lambda)/x!$, $x = 0, 1, \ldots$

Fisher Information MLE:

$$L(\lambda|D) = \prod_{i=1}^{n} p(x_i|\lambda) = \sum_{i=1}^{n} \ln p(x_i|\lambda)$$

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}} \sum_{i=1}^{n} \left[x_i \ln(\lambda) - \lambda - \ln(x_i!) \right]$$

$$= \underset{\lambda}{\operatorname{argmax}} \left[\ln(\lambda)\, n\mu - n\lambda - \sum \ln(x_i!) \right] - 1$$

$\hookrightarrow \mu = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\frac{\partial 1}{\partial \lambda} = \frac{n\mu}{\lambda} - n = 0 \rightarrow \hat{\lambda} = \mu$$

$$\frac{\partial^2 1}{\lambda \lambda^2} = \frac{-n\mu}{\lambda^2} = \frac{-n}{\mu} < 0 \leftarrow \text{maximum!}$$

$$\boxed{\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i} \leftarrow \text{from Fisher}$$

CLT: ← States that given enough samples, data will follow normal distribution $\sim N(\mu, \sigma^2)$ ← This means that

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmax}}\ N(\mu, \sigma^2) = \mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\boxed{\text{from CLT} \rightarrow \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

3) $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$

$Y = X\beta + \varepsilon$, where $X = [\underline{1} \;\; \underline{x}_1 \;\; \cdots \;\; \underline{x}_p]$

Least Squares: $\beta = (X^T X)^{-1} X^T y$

MLE: $\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \; p(Y|D), \quad D = \{x_1, \ldots, x_p\}$

$= \underset{\beta}{\operatorname{argmax}} \prod_{i=1}^{n} p(y_i | x_i)$

$\underline{\varepsilon} = \underline{y} - X\underline{\beta} = N(0, \sigma^2)$

$E[y - X\beta] = 0$

$\operatorname{var}(y - X\beta) = \sigma^2$

$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{y_i - X\beta_i}{\sigma}\right)^2\right)$

$= \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2}\left(\frac{y_i - X\beta_i}{\sigma}\right)^2\right] - ①$

$\frac{\partial ①}{\partial \beta} = \sum_{i=1}^{n} \frac{X^T}{\sigma}\left(\frac{y_i - X\beta_i}{\sigma}\right) = 0$

$= \frac{X^T}{\cancel{\sigma^2}}(\underline{y} - X\beta) = 0 \rightarrow X^T X \beta = X^T y$

$\boxed{\hat{\beta} = (X^T X)^{-1} X^T y}$

$\frac{\partial^2 ①}{\partial \beta^2} = \frac{-X^T X}{\sigma^2} < 0 \leftarrow$ concave down, maximum!

4) $Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon \,, \quad \varepsilon \sim N(0, \sigma^2)$

$Y = X\vec{\beta} + \varepsilon \,, \quad X = [1 \; X_1 \; \cdots \; X_p]$

$\varepsilon = y - X\vec{\beta} \sim N(0, \sigma^2)$

$$P_{Y|X}(Y|x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{y - x_i\beta_i}{\sigma}\right)^2\right)$$

MAP: $$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \left[\prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y - X_i\beta_i}{\sigma}\right)^2\right]\right] \cdot$$

$$\sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left[-\frac{\lambda}{2}\left(\frac{\beta_i}{\sigma}\right)^2\right]$$

$$\hat{\beta} = \underset{\beta}{\operatorname{argmax}} \sum_{i=1}^{p}\left[-\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2}\left(\frac{y - x_i\beta_i}{\sigma}\right)^2\right] +$$

$$\frac{1}{2}\ln(\lambda) - \frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{\lambda}{2}\left(\frac{\beta_i}{\sigma}\right)^2$$

$$= \underset{\beta}{\operatorname{argmin}} \left[\frac{p+1}{2}\ln(2\pi) + (p+1)\ln(\sigma) - \frac{1}{2}\ln(\lambda) + \frac{\lambda}{2}\left(\frac{\beta_i}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{y - x_i\beta_i}{\sigma}\right)^2\right] - 1$$

$$\frac{\partial 1}{\partial \beta} = \frac{\lambda\beta_i}{\cancel{\sigma^2}} - X^T\left(\frac{y - X\beta}{\cancel{\sigma^2}}\right) = 0$$

$$(X^TX + \lambda)\beta = X^Ty$$

$$\boxed{\hat{\beta} = (X^TX + \lambda I)^{-1}X^Ty}$$

λ adds bias to estimate to give a more accurate est. based on probabalistic structure!

5) $\vec{Y} = X\vec{\beta} + \vec{\varepsilon}$, $\vec{\varepsilon} \sim N(0, \sigma^2)$

$$p(\beta_i) = \text{Lap}(0, \sigma^2/\lambda) = \frac{\lambda}{2\sigma^2} \exp\left(-\frac{|\beta_i|}{\sigma^2}\right)$$

$$\varepsilon = y - X\vec{\beta} \sim N(0, \sigma^2)$$

$$p_{Y|X}(Y|x_i) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{y - x_i\beta_i}{\sigma}\right)^2\right)$$

$$\text{MAP: } \hat{\beta} = \underset{\beta}{\text{argmax}} \left[\prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y - x_i\beta_i}{\sigma}\right)^2\right]\right] \cdot \frac{\lambda}{2\sigma^2} \exp\left(-\frac{|\beta_i|}{\sigma^2}\right)$$

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{p} \left[+\frac{1}{2}\ln(2\pi) + \ln(\sigma) + \frac{1}{2}\left(\frac{y - x_i\beta_i}{\sigma}\right)^2\right] + \ln\left(\frac{\lambda}{2\sigma^2}\right) + \frac{|\beta_i|}{\sigma^2}$$

$$\nabla_\beta \hat{\beta} = \begin{cases} \frac{1}{\sigma^2}\left[-X^T(Y_i - X\beta) + 1\right] & \beta_i \geq 0 \\ \frac{1}{\sigma^2}\left[-X^T(Y_i - X\beta) - 1\right] & \text{otherwise} \end{cases} = 0$$

$$\beta_i \geq 0: (X^TX)\beta = X^TY - \frac{1}{\sigma^2} \rightarrow \beta = (X^TX)^{-1}\left(X^TY - \frac{1}{\sigma^2}\right)$$

$$\beta_i < 0: (X^TX)\beta = X^TY + \frac{1}{\sigma^2} \rightarrow \beta = (X^TX)^{-1}\left(X^TY + \frac{1}{\sigma^2}\right)$$

$\nabla^2_\beta \hat{\beta} = X^TX > 0 \leftarrow$ pos def, maximum.

$$\hat{\beta} = \begin{cases} (X^T X)^{-1}\left(X^T Y - \frac{1}{\sigma^2}\right) & \beta_i \geq 0 \\ (X^T X)^{-1}\left(X^T Y + \frac{1}{\sigma^2}\right) & \text{otherwise} \end{cases}$$

I believe this is a different form of bias introduced as a result of the a priori probabalistic structure $p(\beta_i)$. $\therefore$ $\hat{\beta}$ is no longer an unbiased estimator, which is the same as the last problem.

6) $X = U\Sigma V^T$

a. $$\beta^{\wedge R} = \underset{\beta}{\operatorname{argmin}}\ RSS + \lambda \sum_{j=1}^{p} \beta_j^2 = \underset{\beta}{\operatorname{argmin}}\ \|y - X\beta\|^2 + \lambda \|\beta\|^2$$

$$= \underset{\beta}{\operatorname{argmin}}\ (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$$

$$= \underset{\beta}{\operatorname{argmin}} \left(y^T y - (X\beta)^T y - y^T (X\beta) + (X\beta)^T (X\beta) + \lambda \beta^T \beta \right)$$

$$= \underset{\beta}{\operatorname{argmin}} \left(y^T y - \beta^T X^T y - y^T X \beta + \beta^T (X^T X + \lambda I) \beta \right) \sim 1$$

$$\frac{\partial 1}{\partial \beta} = -X^T y - (y^T X)^T + 2\beta (X^T X + \lambda I) = 0$$

$$\cancel{2}\beta (X^T X + \lambda I) = \cancel{2} X^T y$$

$$\hat{\beta}^R = (X^T X + \lambda I)^{-1} X^T y$$

$$\beta^{\wedge R} = ((U\Sigma V^T)^T (U\Sigma V^T) + \lambda I)^{-1} (V \Sigma^T U^T) y$$

$$= (V \Sigma^T U^T U \Sigma V^T + \lambda I)^{-1} (V \Sigma^T U^T) y, \quad \Sigma^T = \Sigma$$

$$= (\Sigma^T \Sigma + \lambda I)^{-1} (V \Sigma U^T) y$$

$$\hat{y} = X\beta^{\wedge R} = U \Sigma V^T (\Sigma^T \Sigma + \lambda I)^{-1} (V \Sigma U^T) y$$

$$= \sum_{j=1}^{p} \sigma_j u_j v_j^T (\sigma_j^2 + \lambda)^{-1} v_j \sigma_j u_j^T y$$

$$= \sum_{j=1}^{p} \sigma_j u_j v_j^T (\sigma_j^2 + \lambda)^{-1} v_j \sigma_j u_j^T y$$

$$= \sum_{j=1}^{p} \frac{\sigma_j^2}{\sigma_j^2 + \lambda} u_j \underbrace{\boxed{v_j^T v_j}}_{= I} u_j^T y$$

$$\boxed{\hat{y} = X \hat{\beta}^{re} = \sum_{j=1}^{p} \frac{\sigma_j^2}{\sigma_j^2 + \lambda} u_j u_j^T y}$$

6. $tr(L_{n \times n}) = \sum_{i=1}^{n} L_{ii}$

$$X(X^T X + \lambda I)^{-1} X^T = U \Sigma V^T (V \Sigma^T U^T U \Sigma V^T + \lambda I)^{-1} V \Sigma U^T$$

$$= U \Sigma V^T (V \Sigma^T \Sigma V^T + \lambda I)^{-1} V^T \Sigma U$$

$$= U \Sigma V^T (\Sigma^T \Sigma \cancel{(V V^T)}^{I} + \lambda I)^{-1} V \Sigma U^T$$

$$= \sum_{i=1}^{p} u_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda} u_i^T$$

$$tr\left(\sum_{i=1}^{p} u_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda} u_i^T \right) = \sum_{j=1}^{n} \sum_{i=1}^{p} \left(u_i \frac{\sigma_i^2}{\sigma_i^2 + \lambda} u_i^T \right)_{jj}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{p} \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda} u_i u_i^T \right)_{jj} = \sum_{j=1}^{n} \frac{\sigma_j^2}{\sigma_j^2 + \lambda} \cancel{\sum_{i=1}^{p} (u_i u_i^T)_{jj}}^{I}$$

$$= \sum_{j=1}^{n} \frac{\sigma_j^2}{\sigma_j^2 + \lambda}$$

$$\boxed{\therefore \text{trace}\left(X(X^T X + \lambda I)^{-1} X^T\right) = \sum_{j=1}^{n} \frac{\sigma_j^2}{\sigma_j^2 + \lambda}}$$