1. (a) $Y = \beta_0 + \beta_1 X + \varepsilon$, ε, X : indep & $\varepsilon \sim N(0, \sigma^2 I)$

$\underline{\beta} = (\beta_0, \beta_1)^T$

$$\underset{\sim}{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$SE(\hat{\beta}_j)^2 = \{\sigma^2 (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}\}_{jj}$$

$$\underset{\sim}{X}^T \underset{\sim}{X} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

$$(\underset{\sim}{X}^T \underset{\sim}{X})^{-1} = \frac{1}{\det(\underset{\sim}{X}^T \underset{\sim}{X})} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} = \frac{1}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$\Rightarrow SE(\hat{\beta}_0)^2 = \{\sigma^2 (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}\}_{00}$$

$$= \frac{\sigma^2 \sum_{i=1}^n x_i^2}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2} = \frac{\sigma^2}{n} \left(\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right)$$

$$= \frac{\sigma^2}{n} \left(\frac{\sum x_i^2 - n\bar{x}^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} - \frac{\sum x_i^2 - n\bar{x}^2}{\sum x_i^2 - n\bar{x}^2} + \frac{\sum x_i^2}{\sum x_i^2 - n\bar{x}^2} \right)$$

$$= \frac{\sigma^2}{n} \left(1 + \frac{n\bar{x}^2}{\sum x_i^2 - n\bar{x}^2} \right) \quad \because \sum x_i^2 - n\bar{x}^2 = \sum (x_i - \bar{x})^2$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} \right)$$

$$SE(\hat{\beta}_1)^2 = \{\sigma^2 (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}\}_{11}$$

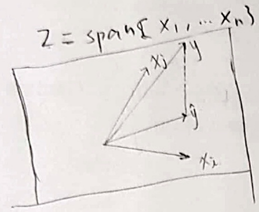
$$= \frac{\sigma^2 n}{n\sum_{i=1}^n x_i^2 - n^2\bar{x}^2}$$

$$= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Scanned by CamScanner

1. (b) Show $Hz = z$, where H : project matrix & $z = Xc$, $c = (c_0, \dots c_p)^T$

$Z = \text{span}\{x_1, \dots x_n\}$



$$Hy = \hat{y} = X(X^TX)^{-1}X^Ty$$

$$HXc = X(X^TX)^{-1}X^TXc = Xc$$

Scanned by CamScanner

2. (a)

$$\hat{\beta} = \arg\max_{\beta} P(y|X,\beta)$$
$$= \arg\max_{\beta} \prod_{i=1}^{n} N(y_n | x_i^T \beta)$$
$$= \arg\max_{\beta} \prod_{i=1}^{n} (2\pi\sigma^2)^{\frac{-1}{2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - x_i^T\beta)^2\right)$$
$$= \arg\max_{\beta} (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2\right)$$
$$= \arg\max_{\beta} -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2$$
$$= \arg\min_{\beta} \sum_{i=1}^{n}(y_i - x_i^T\beta)^2$$

$$L = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - X_i^T\beta)^2 = \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow -2X^Ty + 2X^TX\beta = 0$$
$$\Rightarrow \beta = (X^TX)^{-1}X^Ty.$$

(b)

$$\hat{\beta} = \arg\max_{\beta} P(\beta|y,X)$$
$$= \arg\max_{\beta} \frac{P(y|X,\beta)P(\beta|X)}{P(y|X)}$$
$$= \arg\max_{\beta} P(y|X,\beta)P(\beta|X)$$
$$= \arg\max_{\beta} \log P(y|X,\beta) + \log P(\beta|X)$$
$$= \arg\max_{\beta} -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 - \frac{n}{2}\log(2\pi\gamma^2) - \frac{1}{2\gamma^2}\sum_{i=1}^{n}\beta_i^2$$
$$= \arg\min_{\beta} \sum_{i=1}^{n}(y_i - X_i^T\beta)^2 + \frac{\sigma^2}{\gamma^2}\sum_{i=1}^{n}\beta_i^2$$

$$L = \sum_{i=1}^{n}(y_i - X_i^T\beta)^2 + \frac{\sigma^2}{\gamma^2}\sum_{i=1}^{n}\beta_i^2$$
$$= (y - X\beta)^T(y - X\beta) + \frac{\sigma^2}{\gamma^2}\beta^T\beta$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow (-2X^Ty + 2X^TX\beta) + \frac{\sigma^2}{\gamma^2}2\beta = 0$$
$$\Rightarrow \beta = (X^TX + \frac{\sigma^2}{\gamma^2}I)^{-1}X^Ty.$$

Scanned by CamScanner

2. (c)

$$\text{ML}: \hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2$$

$$\text{MAP}: \hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \frac{\sigma^2}{\tau^2} \sum_{i=1}^{n} \beta_i^2$$

least square sol.

$\Rightarrow$ The prior distribution acts as a regularier in MAP estimation

Scanned by CamScanner

3. $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

(a)

$x_i = 5, 10, 15, 20, 30\,(\%) \rightarrow \bar{x} = 0.16.$

$y_i = \frac{33}{200}, \frac{48}{200}, \frac{62}{200}, \frac{99}{200}, \frac{132}{200} \longrightarrow \bar{y} = 0.374.$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\Rightarrow \hat{\beta}_0 n\bar{x} + \hat{\beta}_1 \sum_i x_i^2 = \sum_i x_i y_i \quad -①$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2\sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\Rightarrow n\hat{\beta}_0 + \hat{\beta}_1 n\bar{x} = n\bar{y}$$

$$\Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad -② \qquad \hat{\beta}_0 = \underline{0.042976}\ \#$$

$$② \rightarrow ① : \ n\bar{x}\bar{y} - n\hat{\beta}_1\bar{x}^2 + \hat{\beta}_1 \sum_i x_i^2 = \sum_i x_i y_i$$

$$\Rightarrow \left(\sum_i x_i^2 - n\bar{x}^2\right)\hat{\beta}_1 = \sum_i x_i y_i - n\bar{x}\bar{y}$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum_i x_i y_i - n\bar{x}\bar{y}}{\sum_i x_i^2 - n\bar{x}^2} = \frac{75.15 - 59.84}{33 - 25.6} = \underline{2.0689}\ \#$$

$$\Rightarrow \underline{\hat{y} = 0.043 + 2.0689x}\ \#$$

$$\Rightarrow \underline{x = 0.25 \quad \Rightarrow \hat{y} = 0.5602}\ \#$$

3. (b)

95%-confidence interval of $\hat{\beta}_0$: $[\hat{\beta}_0 - 2SE(\hat{\beta}_0), \hat{\beta}_0 + 2SE(\hat{\beta}_0)]$

95%-confidence interval of $\hat{\beta}_1$: $[\hat{\beta}_1 - 2SE(\hat{\beta}_1), \hat{\beta}_1 + 2SE(\hat{\beta}_1)]$

where $SE(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]$

$SE(\hat{\beta}_1)^2 = \sigma^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\sigma^2 = \frac{RSS}{n-2}$

RSS:

$\frac{33}{200} - 0.043 - 2.0689 \cdot 0.05 = 0.01855$

$\frac{48}{200} - 0.043 - 2.0689 \cdot 0.10 = -0.00989$

$\frac{62}{200} - 0.043 - 2.0689 \cdot 0.15 = -0.043335$

$\frac{99}{200} - 0.043 - 2.0689 \cdot 0.2 = 0.03822$

$\frac{132}{200} - 0.043 - 2.0689 \cdot 0.3 = -0.00367$

$(0.01855^2 + (-0.00989)^2 + (-0.043335)^2 + (0.03822)^2 + (-0.00367)^2) \times 200 = 0.7588$

$\sigma^2 = 0.7588 / 998 = 7.603 \times 10^{-4}$

$SE(\beta_0)^2 = 7.603 \times 10^{-4} \times \left[\frac{1}{1000} + \frac{0.16^2}{7.4} \right] = 3.39 \times 10^{-6}$

$SE(\beta_1)^2 = 7.603 \times 10^{-4} / 7.4 = 1.027 \times 10^{-4}$

Ans: $\hat{\beta}_0 : 0.043 \pm 2SE(\beta_0) \Rightarrow [0.0393, 0.0467]$

$\hat{\beta}_1 = 2.0689 \pm 2SE(\beta_1) \Rightarrow [2.0487, 2.0891]$

3(c)

$$l(B_0, B_1) = \prod_{i=1}^{n} P(Y = y_i \mid X = x_i ; B_0, B_1) = \prod_{y_i=1} (P(x_i ; B_0, B_1)) \prod_{y_i=0} (1 - P(x_i ; B_0, B_1))$$

$$\ln l(B_0, B_1) = \sum_{i=1}^{n} (y_i \cdot \ln P(x_i ; B_0, B_1) + (1-y_i) \ln P(1 - P(x_i ; B_0, B_1)))$$

$$= \sum_{i=1}^{n} \left[y_i (B_0 + B_1 x_i) - \ln (1 + e^{B_0 + B_1 x_i}) \right]$$

$$= - \sum_{i=1}^{n} \ln (1 + e^{-(3y_i - 1) B^T x_i}) \triangleq J(B_0, B_1)$$

$$= -33 \ln (1 + e^{-(B_0 + 5B_1)}) - (200 - 31) \ln (1 + e^{(B_0 + B_1 x)})$$

$$- 48 \ln (1 + e^{-(B_0 + 10 B_1)}) - (200 - 48) \ln (1 + e^{-(B_0 + 10 B_1)})$$

$$- 62 \ln (1 + e^{-(B_0 + 15 B_1)}) - (200 - 62) \ln (1 + e^{-(B_0 + 15 B_1)})$$

$$- 99 \ln (1 + e^{-(B_0 + 15 B_1)}) - (200 - 99) \ln (1 + e^{-(B_0 + 20 B_1)})$$

$$- 132 \ln (1 + e^{-(B_0 + 30 B_1)}) - (200 - 132) \ln (1 + e^{-(B_0 + 30 B_1)})$$

use Gradient Descent $B(k+1) = B(k) - \eta \cdot \nabla J(B(k))$ #

4.

Drop a coefficient $\Rightarrow q = 1$

$$F = \frac{[RSS(\hat{\hat{\beta}}_{1:p-q}) - RSS(\hat{\beta})]/q}{RSS(\hat{\beta})/n-p-1} = \frac{[RSS(\hat{\hat{\beta}}_{1:p-1}) - RSS(\hat{\beta})]}{RSS(\hat{\beta})/n-p-1} \Rightarrow \begin{cases} RSS(\hat{\hat{\beta}}_{1:p-1}) - RSS(\hat{\beta}) \sim \sigma^2 \chi^2_1 \\ RSS(\hat{\beta}) \sim \sigma^2 \chi^2_{n-p-1} \end{cases}$$

$$t = \frac{\hat{\beta}_j}{\hat{\sigma}^2 \{(X^T X)^{-1}\}_{jj}} = \frac{\hat{\beta}_j}{\sqrt{\frac{\hat{\sigma}^2 (n-p-1) \{(X^T X)^{-1}\}_{jj}}{n-p-1}}} = \frac{\frac{\sigma \hat{\beta}_j}{\underbrace{\sqrt{\sigma^2 \{(X^T X)^{-1}\}_{jj}}}_{Var(\hat{\beta}_j)}}}{\sqrt{\frac{\hat{\sigma}^2 (n-p-1)}{n-p-1}}} \Rightarrow \begin{cases} \frac{\sigma^2 \hat{\beta}_j^2}{Var(\hat{\beta}_j)} \sim \sigma^2 \chi^2_1 \\ \hat{\sigma}^2 (n-p-1) \sim \sigma^2 \chi^2_{n-p-1} \end{cases}$$

$\because$ The distributions of F-statistic and the square of t-score are equal

If $p = 1$

$$\begin{aligned} RSS(\hat{\hat{\beta}}) - RSS(\hat{\beta}) &= \sum (y_i - \bar{y})^2 - \sum (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i)^2 \\ &= \cancel{\sum (y_i - \bar{y})^2} - \left(\cancel{\sum (y_i - \bar{y})^2} + \sum \hat{\beta}_1^2 (x_i - \bar{x})^2 - 2 \sum (y_i - \bar{y}) \hat{\beta}_1 (x_i - \bar{x}) \right) \\ &= \hat{\beta}_1^2 \sum (x_i - \bar{x})^2 \end{aligned}$$

$$\Rightarrow F = \frac{\hat{\beta}_1 \sum (x_i - \bar{x})^2}{RSS(\hat{\beta})/n-2} = \frac{\frac{\sigma \hat{\beta}_j}{\sqrt{\sigma^2 \{(X^T X)^{-1}\}_{jj}}}}{\sqrt{\frac{\hat{\sigma}^2 (n-p-1)}{n-p-1}}} = t^2$$

5. (a) If $X : U(0,1)$, then $\frac{0.65 - 0.55}{1-0} = 10\%$.

(b) $10\% \cdot 10\% = 1\%$

(c) $10\%^{100} = 10^{-100}$

(e) For $p=1$, side $= 0.1$

For $p=2$, side $= 0.1^{\frac{1}{2}} = 0.316$

For $p=100$, side $= 0.1^{\frac{1}{100}} = 0.977$

when p is too high, to use on average 10% of the observations would mean that we would need to include almost the entire range of each individual feature.

6. (a)

$$Q = \sum_{i=1}^{n} (y_i - (B_0 + B_1 x_i))^2 \quad , \quad y_i = f(x_i)$$

$$\Rightarrow \frac{\partial Q}{\partial B_0} = -2\sum (y_i - (B_0 + B_1 x_i)) = 0$$

$$\frac{\partial Q}{\partial B_1} = -2\sum x_i (y_i - (B_0 + B_1 x_i)) = 0$$

$$\Rightarrow \sum y_i = n B_0 + B_1 \sum x_i . \quad \sum x_i y_i = \sum \left(\beta_0 + \beta_1 x_i\right)$$

$$\sum x_i y_i = \sum x_i (B_0 + B_1 x_i)$$

$$\Longrightarrow \hat{B_1} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \quad \hat{B_0} = \bar{y} - \hat{B_1}\bar{x} . \quad \bar{y} = \frac{1}{n}\sum y_i \quad \bar{x} = \frac{1}{n}\sum x_i$$

$$\Rightarrow C_{xx}^{(n)} = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

$$= \frac{1}{n}\sum x_i^2 - \frac{(\sum x_i)^2}{n^2}$$

$$C_{xy}^{(n)} = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y} = \frac{1}{n}\sum x_i y_i - \frac{1}{n^2}\sum x_i \sum y_i$$

$$\Rightarrow \hat{\beta}_1 = \frac{n^2 C_{xy}}{n^2 C_{xx}} = \frac{C_{xy}}{C_{xx}} , \quad \hat{\beta}_0 = \hat{y} - \hat{\beta}_1^{(n)} \bar{x}^{(n)}$$

6.

(b)(c)

$$\bar{x}^{(n+1)} = \frac{1}{n+1} \cdot \sum_{i=1}^{n+1} x_i = \frac{1}{n+1}\left(\sum_{i=1}^{n} x_i + x_{n+1}\right)$$

$$= \frac{1}{n+1}\left(\frac{-\sum_{i=1}^{n} x_i}{n} + \frac{\left(\sum_{i=1}^{n} x_i\right)\cdot(n+1)}{n} + x_{n+1} \right)$$

$$= \bar{x}^{(n)} + \frac{1}{n+1}\left(x_{n+1} - \bar{x}^{(n)}\right)$$

as above, $\bar{y}^{(n+1)} = \bar{y}^{(n)} + \frac{1}{n+1}\left(x_{n+1} - \bar{y}^{(n)}\right)$

---

next page.

$$C_{xx}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}^{(n)})^2$$

$$C_{xy}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x}^{(n)})(y_i - \bar{y}^{(n)})$$

$$\longrightarrow C_{xy}^{(n+1)} = \frac{1}{n+1} \cdot \sum_{i=1}^{n+1} (x_i - \bar{x}^{(n+1)})(y_i - \bar{y}^{(n+1)})$$

$$= \frac{1}{n+1} \sum_{i=1}^{n+1} (x_i y_i + \bar{x}^{(n+1)} \bar{y}^{(n+1)} - x_i \bar{y}^{(n+1)} - \bar{x}^{(n+1)} y_i)$$

from above $= \left[\frac{1}{n+1} \sum_{i=1}^{n+1} (x_i y_i)\right] + \bar{x}^{(n+1)} \bar{y}^{(n+1)} - 2\bar{x}^{(n+1)} \bar{y}^{(n+1)}$

$$= \frac{1}{n+1}\sum_{i=1}^{n+1} (x_i y_i) - \bar{x}^{(n+1)} \bar{y}^{(n+1)}$$

$$= \frac{x_{n+1} y_{n+1} - (n+1)\bar{x}^{(n+1)} \bar{y}^{(n+1)}}{n+1} + \frac{1}{n+1}\left(\sum_{i=1}^{n} x_i y_i\right)$$

$$= \frac{1}{n+1}\left[x_{n+1} y_{n+1} + n C_{xy}^{(n)} + n \bar{x}^{(n)} \bar{y}^{(n)} - (n+1)\bar{x}^{(n+1)} \bar{y}^{(n+1)}\right]$$

as above $C_{xx}^{(n+1)} = \frac{1}{n+1}\left[x_{n+1}^2 + n C_{xx}^{(n)} + n \bar{x}^{(n)2} - (n+1)\bar{x}^{(n+1)2}\right]$

(d) When new data x_{n+1} arrives, we can obtain the optimal solution by the $\bar{x}^{(n)}$, $\bar{y}^{(n)}$, $C_{xy}^{(n)}$, $C_{xx}^{(n)}$, instead of computing the whole n+1 data, in the purpose of saving computation and memory cost.