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

$\beta = (\beta_0, \beta_1)^T$

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

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

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

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

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

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

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

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

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

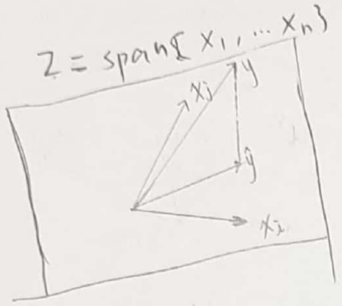
$$SE(\hat{\beta}_1) = \{\sigma^2 (X^T X)^{-1}\}_{11}$$

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

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

Scanned by CamScanner

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



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

$$H = X(X^TX)^{-1}X^T$$

2. (a)

$$\hat{\beta} = \max_{\beta} \, p(y|X,\beta,\sigma^2)$$

$$p(y|X,\beta,\sigma^2) = \prod_{i=1}^{n} N(y_n|x_i^T\beta, \sigma^2)$$

$$L = \ln p(y|X,\beta,\sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow \quad -\frac{1}{2\sigma^2}\cdot 2\sum_{i=1}^{n}(y_i - x_i^T\beta)(-x_i^T) = 0.$$

$$\Rightarrow \beta = (X^TX)^{-1}X^Ty.$$

(b)

$$\hat{\beta} = \max_{\beta} \, p(\beta|y,X) = \frac{p(y|X,\beta)\,p(\beta|X)}{p(y|X)}$$

$$L = \ln p(\beta|y,X) = \ln p(y|X,\beta) + \ln p(\beta|X) - \ln p(y|X)$$

$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 - \frac{n}{2}\ln(2\pi r^2) - \frac{1}{2r^2}\beta^T\beta$$

$$\frac{\partial L}{\partial \beta} = 0 = \frac{\cancel{2}}{\cancel{2}\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)(-x_i^T) - \frac{\cancel{2}}{\cancel{2}r^2}\sum_{j=0}^{p}\beta_j$$

$$\Rightarrow \beta = \left(\frac{\sigma^2}{r^2}I + X^TX\right)^{-1}X^Ty$$

Scanned by CamScanner

2.(c)

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

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

least square sol.

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

3 (a) linear regression

$\bar{x} = 0.16$

$\bar{y} = 0.396$

$y = \hat{\beta}_0 + \hat{\beta}_1 X$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{1000} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{1000} (x_i - \bar{x})^2} = \frac{16.99}{7.4} = 2.296$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 0.02865$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X = 0.02865 + 2.296 X$$

$\Rightarrow$ $X = 0.25 \Rightarrow \hat{y} = 0.603$

3 (b) $l(\beta_0, \beta_1) = \prod_{i=1}^{n} P(Y = y_i \mid X = x_i; \beta_0, \beta_1) = \prod_{y_i=1} P(x_i; \beta_0, \beta_1) \prod_{y_i=0} (1 - P(x_i; \beta_0, \beta_1))$

$$\ln l(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i \ln P(x_i; \beta_0, \beta_1)) + (1 - y_i) \ln P(1 - P(x_i; \beta_0, \beta_1)))$$

$$= \sum_{i=1}^{n} \left[y_i (\beta_0 + \beta_1 x_i) - \ln(1 + e^{\beta_0 + \beta_1 X}) \right]$$

$$= -\sum_{i=1}^{n} \ln(1 + e^{-(2y_i - 1)\beta^T x_i}) \triangleq J(\beta_0, \beta_1)$$

$$= -31 \ln(1 + e^{-(\beta_0 + 5\beta_1)}) - (200 - 31) \ln(1 + e^{(\beta_0 + 5\beta_1)})$$
$$-52 \ln(1 + e^{-(\beta_0 + 10\beta_1)}) - (200 - 52) \ln(1 + e^{(\beta_0 + 10\beta_1)})$$
$$-68 \ln(1 + e^{-(\beta_0 + 15\beta_1)}) - (200 - 68) \ln(1 + e^{(\beta_0 + 15\beta_1)})$$
$$-101 \ln(1 + e^{-(\beta_0 + 20\beta_1)}) - (200 - 101) \ln(1 + e^{(\beta_0 + 20\beta_1)})$$
$$-144 \ln(1 + e^{-(\beta_0 + 30\beta_1)}) - (200 - 144) \ln(1 + e^{(\beta_0 + 30\beta_1)})$$

Use Gradient Descent $\beta(k+1) = \beta(k) - \eta \nabla J(\beta(k))$

Scanned by CamScanner

3.(c)

$$P(Y=1|X=x) = \frac{P(X=x|Y=1)P(Y=1)}{P(X=x)} \triangleq P_1(x)$$

$$P_1(x) = \hat{\delta}_1(x) = x\frac{\hat{\mu}_1}{\hat{\sigma}^2} - \frac{\hat{\mu}_1^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_1)$$

$$\pi_1 = \frac{396}{1000} \doteq 0.396$$

$$\hat{\mu}_1 = \frac{1}{n_1}\sum_{y_i=1} x_i = \frac{1}{396}(31\cdot 5 + 52\cdot 10 + 68\cdot 15 + 101\cdot 20 + 144\cdot 30) = 20.2904$$

$$\hat{\mu}_0 = \frac{1}{n_0}\sum_{y_i=0} x_i = \frac{1}{604}(169\cdot 5 + 148\cdot 10 + 132\cdot 15 + 99\cdot 20 + 56\cdot 30) = 13.1871$$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{k=0}^{1}\sum_{y_i=k}(x_i - \hat{\mu}_k)^2$$

$$= \frac{1}{1000-2}\Big(169\cdot(5-13.1871)^2 + 148(10-13.1871)^2 + 132(15-13.1871)^2$$
$$+ 99(20-13.1871)^2 + 56(30-13.1871)^2 + 31(5-20.2904)^2$$
$$+ 52(10-20.2904)^2 + 68(15-20.2904)^2 + 101(20-20.2904)^2$$
$$+ 144(30-20.2904)^2\Big)$$

$$= 62.056$$

$$P_1(x) = \frac{\pi_1 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\hat{\mu}_1)^2\right)}{\sum_{l=0}^{1} \pi_l \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\hat{\mu}_l)^2\right)}$$

$$x = 25 \Rightarrow P_1(25) = 0.628$$

Scanned by CamScanner

4.

$$R^2 = \frac{TSS-RSS}{TSS} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{j=1}^{n} y_j^2}$$

$$TSS = \sum_{j=1}^{n}(y_j - \bar{y})^2 = \sum_{j=1}^{n} y_j^2 \quad \because \bar{y} = 0$$

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2 \qquad \hat{\beta}_0 = 0 \because \bar{x} = \bar{y} = 0$$

$$= \sum_{i=1}^{n}\left(y_i - \left(\frac{\sum_{k=1}^{n} x_k y_k}{\sum_{k=1}^{n} x_k^2}\right) x_i\right)^2$$

$$Cor(X,Y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}$$

$$\Rightarrow R^2 = [Cor(X,Y)]^2$$

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.

Scanned by CamScanner