# 1 Logistic regression

a)

$$P(Y = y_i|X = x_i) = \sigma(b + \omega^T x_i)^{y_i}[1 - \sigma(b + \omega^T x_i)]^{1-y_i}$$

Likehood $L(\omega) = \prod_{i=1}^{N} \sigma(b + \omega^T x_i)^{y_i}[1 - \sigma(b + \omega^T x_i)]^{1-y_i}$

$$\log L(\omega) = \sum_{i=1}^{N} y_i \log \sigma(b + \omega^T x_i) + (1 - y_i) \log[1 - \sigma(b + \omega^T x_i)]$$

Negative log like hood is

$$\epsilon(\omega) = -\sum_{i=1}^{N}[y_i \log \sigma(b + \omega^T x_i) + (1 - y_i) \log[1 - \sigma(b + \omega^T x_i)]]$$

b)
for $f = y_i \log \sigma(b + \omega^T x_i)$ let $A = \omega^T x_i$, $B = \sigma(A)$

$$\frac{df}{d\omega^T} = y_i \frac{d \log B}{dB} \frac{d\omega(A)}{dA} \frac{d(b + \omega^T x_i)}{d\omega^T}$$

$$= \frac{1}{\sigma(b + \omega^T x_i)} \sigma(b + \omega^T x_i)[1 - \sigma(b + \omega^T x_i)]x_i$$

$$= y_i[1 - \sigma(b + \omega^T x_i)]x_i$$

for $g = (1 - y_i) \log[1 - \sigma(b + \omega^T x_i)]$ let $A = \omega^T x_i$, $B = \sigma(A)$, $C = 1 - B$

$$\frac{dg}{d\omega^T} = (1 - y_i) \frac{d \log C}{dC} \frac{d(1 - B)}{dB} \frac{d\sigma(A)}{dA} \frac{d(b + \omega^T x_i)}{d\omega^T}$$

$$= -(1 - y_i)\sigma(b + \omega^T x_i)x_i$$

$$\frac{\partial \epsilon(\omega)}{\partial \omega} = \sum_i [\sigma(b + \omega^T x_i) - y_i]x_i$$

$$\omega^{(t+1)} < -w^{(t)} - \eta \sum_i [\sigma(b + \omega^T x_i) - y_i]x_i$$

c)
log-like hood:

$$\log P(D) = \sum_n \log P(Y_n|X_n)$$

$$= \sum_n \log \prod_{k=1}^{K} P(C_k|X_n)^{y_{nk}}$$

$$= \sum_n \sum_k y_{nk} \log P(C_k|X_n)$$

$$= \sum_n \sum_k y_{nk} \log \frac{\exp(\omega_k^T x)}{1 + \sum_1^k \exp(\omega_t^T x)}$$

negative $\iota(\omega_1, \cdots, \omega_K) = -\sum_n \sum_k y_{nk} \log \frac{\exp(\omega_k^T x)}{1+\sum_1^k \exp(\omega_t^T x)}$

d)

$$\frac{\partial \iota}{\partial \omega} = -\sum_n \sum_i y_{ni} \frac{d \log \mu}{d\mu} \frac{d[\frac{\exp(\omega_i^T x)}{1+\sum_1^K \exp(\omega_t^T x)}]}{d\omega_i}$$

$$= -\sum_n \sum_i y_{ni} \frac{1 + \sum_1^K \exp(\omega_t^T x)}{\exp(\omega_i^T x)} \times \frac{x \exp(\omega_i^T x)(1 + \sum_1^K \exp(\omega_t^T x)) - x \exp(\omega_i^T x) \exp(\omega_i^T x)}{1 + \sum_1^K \exp(\omega_t^T x)}$$

$$= -\sum_n \sum_i y_{ni} \frac{x(1 + \sum_1^K \exp(\omega_t^T x) - \exp(\omega_i^T x))}{1 + \sum_1^K \exp(\omega_t^T x)}$$

$$= -\sum_n \sum_i y_{ni} x (\frac{e^{w_i^T x}}{1 + \sum_1^K e^{w_t^T x}} - 1)$$

# 2 linear regression

(a)

$$\log P(D) = \sum_{n:y_n=1} [\log(P_1 \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_n-\mu_1)^2}{2\sigma_1^2}})] + \sum_{n:y_n=2} [\log(P_2 \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_n-\mu_2)^2}{2\sigma_2^2}})]$$

$$= \sum_{n:y_n=1} [-\frac{(x_n - \mu_1)^2}{2\sigma_1^2} + \log P_1 - \log(\sqrt{2\pi}\sigma_1)] + \sum_{n:y_n=2} [-\frac{(x_n - \mu_2)^2}{2\sigma_2^2} + \log(1-P_1) - \log(\sqrt{2\pi}\sigma_2)]$$

$$\frac{\partial \log P(D)}{\partial P_1} = \sum_{n:y_n=1} \frac{1}{P_1} - \sum_{n:y_n=2} \frac{1}{1 - P_1} = 0$$

$$\sum_{n:y_n=1} (1 - P_1) = \sum_{n:y_n=2} P_1$$

$$nP_1 = \sum_{n:y_n=1} 1$$

$$P_1^* = \frac{\sum_{i=1}^n I(y_i = 1)}{n}$$

$$P_2^* = \frac{\sum_{i=1}^n I(y_i = 2)}{n}$$

$$\frac{\partial \log P(D)}{\partial \mu_1} = \sum_{n:y_n=1} \frac{x_n - \mu_1}{\sigma^2} = 0$$
$$\sum_{n:y_n=1} x_n = \sum_{n:y_n=1} \mu_1$$

$$\mu_1^* = \frac{\sum_{i=1}^n I(y_i = 1)x_i}{\sum_{i=1}^n I(y_i = 1)}$$
$$\mu_2^* = \frac{\sum_{i=1}^n I(y_i = 2)x_i}{\sum_{i=1}^n I(y_i = 2)}$$

$$\frac{\partial \log P(D)}{\partial \sigma_1^2} = \sum_{n:y_n=1} [\frac{(x_n - \mu_1)^2}{\sigma_1^3} - \frac{1}{\sigma_1}] = 0$$
$$= \sum_{n:y_n=1} [\frac{(x_n - \mu_1)^2}{\sigma_1^2} - 1] = 0$$
$$\frac{\sum_{i=1}^n I(y_i = 1)(x_i - \mu_1)^2}{\sigma_1^2} = \sum_{i=1}^n I y_i = 1$$

$$\sigma_1^{*2} = \frac{\sum_{i=1}^n I(y_i = 1)(x_i - \mu_1)^2}{\sum_{i=1}^n I(y_i = 1)}$$
$$\sigma_2^{*2} = \frac{\sum_{i=1}^n I(y_i = 2)(x_i - \mu_2)^2}{\sum_{i=1}^n I(y_i = 2)}$$

(b)

$$P(y = c_k|x)$$
$$= P_{c_k} \exp[\mu_{c_k}^T \sum^{-1} x - \frac{1}{2} x^T \sum^{-1} x - \frac{1}{2} \mu_{c_k}^T \sum^{-1} \mu_{c_k}]$$
$$= exp[\mu_{c_k}^T \sum^{-1} x - \frac{1}{2} \mu_{c_k}^T \sum^{-1} \mu_{c_k} + \log P_{c_k}] \exp[-\frac{1}{2} x^T \sum^{-1} x]$$

$$\gamma_{c_k} = -\frac{1}{2}\mu_{c_k}^T \sum^{-1} \mu_{c_k} + \log P_{c_k}$$

$$\beta_{c_k} = \sum^{-1} \mu_{c_k}$$

$$P(y = c_1|x) = \frac{e^{(}\beta_{c_1}^T x + \gamma_{c_1})}{e^{(}\beta_{c_1}^T x + \gamma_{c_1}) + e^{(}\beta_{c_2}^T x + \gamma_{c_2})}$$

$$P(y = c_1|x) = \frac{1}{1 + e^{[}(\beta_{c_2} - \beta_{c_1})^T x + (\gamma_{c_2} - \gamma_{c_1})]}$$

$$\gamma_{c_2} - \gamma_{c_1} = -\frac{1}{2}(\mu_{c_2} - \mu_{c_1})^T \sum^{-1} (\mu_{c_2} + \mu_{c_1}) + \log(\frac{P_{c_1}}{P_{c_2}})$$

$$\omega = \beta_{c_2} - \beta_{c_1} = \sum^{-1} (\mu_{c_2} - \mu c_1)$$

$$x_0 = \frac{1}{2}(\mu_{c_2} + \mu_{c_1}) - (\mu_{c_2} - \mu_{c_1}) \frac{\log \frac{P_{c_2}}{P_{c_1}}}{(\mu_{c_2} - \mu_{c_1})^T \sum^{-1} (\mu_{c_2} - \mu_{c_1})}$$

$$P(y = c_1|x) = \frac{1}{1 + e^{-\omega^T (x - x_0)}}$$