HOMEWORK # 2 | MAX SOKOLICH

QUESTION # 1 : prove $\frac{\partial LL(\beta)}{\partial \beta} = \sum_{i=1}^{n} \bar{x}_i (y_i - p(\bar{x}_i, \beta))$

where $NLL(\beta) = \sum_{i=1}^{n} y_i \log(p) + (1-y_i)\log(1-p)$

$$\left(p = \frac{1}{1+e^{-\beta^T x_i}} \right)$$

$\hookrightarrow \log(p) = -\log(1+e^{-\beta^T x_i})$

$\hookrightarrow \log(1-p) = \log(e^{-\beta^T x_i}) - \log(1+e^{-\beta^T x_i})$

$$NLL\beta = \sum_{i=1}^{n} \left(-y_i \log(1+e^{-\beta^T x_i}) + (1-y_i)\log(e^{-\beta^T x_i}) - (1-y_i)\left(\log(1+e^{-\beta^T x_i})\right) \right)$$

$$= \sum_{i=1}^{n} \left(y_i \beta^T x_i - \log(1+e^{-\beta^T x_i}) - \beta^T x_i \right)$$

$$\frac{\partial}{\partial \beta}\left(y_i \beta^T x_i \right) = x_i y_i$$

$$\frac{\partial}{\partial \beta}\left(\log(1+e^{-\beta^T x_i}) \right) = (1-p)(-x_i)$$

$$\frac{\partial}{\partial \beta}\left(-\beta^T x_i \right) = -x_i$$

Thus $$\frac{\partial LL\beta}{\partial \beta} = \sum_{i=1}^{n} \left[x_i y_i - (1-p)(-x_i) - (x_i) \right]$$

$\hookrightarrow$

$$\boxed{\frac{\partial NLL(\beta)}{\partial \beta} = \sum x_i (y_i - p)}$$

QUESTION # 2

a) Prove $LL(\beta)$ is concave

$$\frac{\partial LL(\beta)}{\partial \beta} = \sum_{i=1}^{n} x_i (y_i - p)$$

$$\frac{\partial^2 LL(\beta)}{\partial \beta^2} = -\sum_{i=1}^{n} x_i x_i^T \left(p(1-p)\right) \longrightarrow \text{hessian matrix}$$

- $p(1-p)$ is always greater than zero $(0-1)$
- $x_i x_i^T$ are scalars
- therefore $\frac{\partial^2 LL(\beta)}{\partial \beta^2}$ is non positive because of the negative sign and thus negative semidefinite aka concave

b) complete separation of data results in concavity breaking down

↳ all we showed above was uniqueness of minimum but necessarily existance

existance of minimum β

QUESTION #3

let $\sigma(a) = \frac{1}{1+e^{-a}}$

a) use $\sigma(a)$ to formulate $\frac{\partial(\sigma(a))}{\partial a}$

$$\frac{\partial}{\partial a}\left(\frac{1}{1+e^{-a}}\right) = \frac{\partial}{\partial a}\left((1+e^{-a})^{-1}\right) = -(1+e^{-a})^{-2}\frac{\partial}{\partial x}(1+e^{-a})$$

$$= (1+e^{-a})^{-2}(e^{-a}) = \frac{e^{-a}}{(1+e^{-a})^2}$$

$$\boxed{\frac{\partial}{\partial a}(\sigma(a)) = \left(\frac{1}{1+e^{-a}}\right)\left(1 - \frac{1}{1+e^{-a}}\right) = \sigma(a)(1-\sigma(a))}$$

b) show $1-\sigma(a) = \sigma(-a)$

$$1 - \frac{1}{\cancel{1+e^{-a}}} = \frac{1}{\cancel{1+e^{a}}}$$

$$\sigma(a) = \frac{1}{1+e^{-a}}$$

$$= \frac{e^{a}}{e^{a}+1}$$

$$= \frac{e^{a}+1-1}{e^{a}+1}$$

$$= \frac{e^{a}+1}{e^{a}+1} - \frac{1}{e^{a}+1}$$

$$= \boxed{1 - \sigma(-a)}$$

QUESTION #4

suppose $\hat{Y} = X(X^TX)^{-1}X^TY = X\hat{B}$

show LDA using $\hat{Y}$ instead of X is identical to LDA in original space.

need to show new μ, Σ, π are the same

π in $\delta_K(x) = x^T\Sigma^{-1}\mu_K - \frac{1}{2}\mu_K^T\Sigma^{-1}\mu_K + \log \pi_K$

where $\pi_K = N_K/N$

$\mu_K = \sum_{g_i=K} x_i/N_K$

$\hat{\Sigma} = \sum\sum (x_i - \hat{\mu}_K)(x_i - \hat{\mu}_K)^T/(N-K)$

step 1: π_K new

$\underline{\pi_K = \pi_{K\,new}}$ since classification does not change

step 2: μ_K new

$\hat{\mu}_{K\,new} = \sum \frac{x_i B^T}{N_K} = \underline{B^T\hat{\mu}_K}$

step 3: Σ_{new}

$\Sigma_{new} = \sum\sum (x_iB^T - \hat{\mu}_K)(x_iB^T - \mu_K)^T/(N-K)$

$= \frac{1}{N-K}\sum\sum B^T(x_i - \mu_K)(x_i - \mu_K)^T B$

$= \underline{B^T\hat{\Sigma}B}$

new function is: $\delta_{K\,new} = (B^Tx)^T(\Sigma_{new})^{-1}\mu_{K\,new} - \frac{1}{2}(\mu_{K\,new})^T(\Sigma_{new})^{-1}\mu_{K\,new} + \log \pi_{K\,new}$

which $= \boxed{x^T\Sigma^{-1}\mu_K - \frac{1}{2}\mu_K^T\Sigma^{-1}\mu_K + \log \pi_K}$

accidentally answered | 4.3 Before

QUESTION 4.5

$$l(\beta) = \sum_{i=1}^{N} \left[y_i \left(\beta_0 + \beta_1 x_0 + \beta_1 (x_i - x_0) \right) - \log\left(1 + e^{\beta_0 + \beta_1 x_0 + \beta_1 (x_i - x_0)}\right) \right]$$

choose $\beta_0 = -\beta_1 x_0$

$$l(\beta) = \underbrace{\sum_{i: x_i = x_0} \left[-\log\left(1 + e^{\beta_1 (x_i - x_0)}\right) \right]}_{\text{vanishes}} + \underbrace{\sum \left[\beta_1 (x_i - x_0) - \log\left(1 + e^{\beta_1 (x_i - x_0)}\right) \right]}_{\infty}$$

as $\beta_1 \to \infty$, thus max and likelihood will never be achieved

S_1, S_2 subsets of X

a) $$l(\beta) = \sum_{i=1}^{N} \left[y_i \beta^T x_i - \log\left(1 + e^{\beta^T x_i}\right) \right]$$

$$= \sum_{i \in S_1} \beta^T x_i - \log\left(1 + e^{\beta^T x_i}\right) + \sum_{i \in S_2} \left[-\log\left(1 + e^{\beta^T x_i}\right) \right]$$

for $p=1$ $\beta^{new} \leftarrow q \beta^{old}$ with $q \to \infty \Rightarrow l(\beta) \to +\infty$

b) $$l(\beta) = \sum_{k=1}^{K-1} \sum \left[\beta_k^T x_i - \log\left(1 + \sum_{l=1}^{K-1} e^{\beta_l^T x_i}\right) \right] +$$

$$\sum \left[-\log\left(1 + \sum_{l=1}^{K-1} e^{\beta_l^T x_i}\right) \right]$$

there exists β_K for $K = 1 \ldots K$ such that $\beta_K^T x > 0$ for $x \in S_k$

thus, increasing β_K in (a), $l(\beta) \to +\infty$.