CS 146 - Homework 6

Zooey Nguyen

zooeyn@ucla.edu

May 24, 2021

Question 1.

Sigmoid form of conditional probability.

$$
\begin{aligned}
P(C_0|x) &= \frac{P(x|C_0)P(C_0)}{P(x|C_0)P(C_0) + P(x|C_1)P(C_1)} \\
P(C_0|x) &= \frac{1}{1 + \frac{P(x|C_1)P(C_1)}{P(x|C_0)P(C_0)}} \\
P(C_0|x) &= \frac{1}{1 + \exp\left\{\ln \frac{P(x|C_1)P(C_1)}{P(x|C_0)P(C_0)}\right\}} \\
P(C_0|x) &= \frac{1}{1 + \exp\{\ln P(x|C_1)P(C_1) - \ln P(x|C_0)P(C_0)\}} \\
P(C_0|x) &= \frac{1}{1 + \exp\{-(\ln P(x|C_0)P(C_0) - \ln P(x|C_1)P(C_1))\}} \\
P(C_0|x) &= \boxed{\frac{1}{1 + \exp\{-a\}}} \\
a &= \boxed{\ln \frac{P(x|C_0)P(C_0)}{P(x|C_1)P(C_1)}}
\end{aligned}
$$

Linear form of a.

$$
\begin{aligned}
a &= \ln \frac{P(x|C_0)P(C_0)}{P(x|C_1)P(C_1)} \\
a &= (x - \mu_0)^T \Sigma^{-1} (x - \mu_0) P(C_0) - (x - \mu_1)^T \Sigma^{-1} (x - \mu_1) P(C_1) \\
a &= (x^T - \mu_0^T) \Sigma^{-1} (x - \mu_0) P(C_0) - (x^T - \mu_1^T) \Sigma^{-1} (x - \mu_1) P(C_1) \\
a &= [x^T \Sigma^{-1} x + \mu_0^T \Sigma^{-1} \mu_0] P(C_0) - [x^T \Sigma^{-1} x + \mu_1^T \Sigma^{-1} \mu_1] P(C_1) \\
a &= \left[x^T \Sigma^{-1} P(C_0) - x^T \Sigma^{-1} P(C_1)\right] x + \left[\mu_0^T \Sigma^{-1} \mu_0 P(C_0) - \mu_1^T \Sigma^{-1} \mu_1] P(C_1)\right] \\
a &= \boxed{w^T x + b} \\
w &= \left[x^T \Sigma^{-1} P(C_0) - x^T \Sigma^{-1} P(C_1)\right]^T \\
w &= \left[x^T \Sigma^{-1} P(C_0)\right]^T - \left[x^T \Sigma^{-1} P(C_1)\right]^T \\
w &= (P(C_0) \Sigma x - P(C_1) \Sigma x) \\
w &= \boxed{[P(C_0) - P(C_1)] \Sigma x} \\
b &= \boxed{\mu_0^T \Sigma^{-1} \mu_0 P(C_0) - \mu_1^T \Sigma^{-1} \mu_1 P(C_1)}
\end{aligned}
$$

Linear form of a with different covariance matrices. Note that this is the exact same as before but we are adding the following term to a.

$$
\begin{aligned}
a &= \ln \frac{P(x|C_0)P(C_0)}{P(x|C_1)P(C_1)} \\
a &= \ln \frac{|\Sigma_1|^{1/2}}{|\Sigma_0|^{1/2}} \left[(x-\mu_0)^T \Sigma_0^{-1}(x-\mu_0)P(C_0) - (x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)P(C_1)\right] \\
a &= \ln \frac{|\Sigma_1|^{1/2}}{|\Sigma_0|^{1/2}} \left[(x^T \Sigma_0^{-1} x + \mu_0^T \Sigma_0^{-1}\mu_0)P(C_0) - (x^T \Sigma_1^{-1} x + \mu_1^T \Sigma_1^{-1}\mu_1)P(C_1)\right] \\
a &= \ln \frac{|\Sigma_1|^{1/2}}{|\Sigma_0|^{1/2}} \left[(P(C_0)x^T \Sigma_0^{-1} - P(C_1)x^T \Sigma_1^{-1})x + (\mu_0^T \Sigma_0^{-1}\mu_0 - \mu_1^T \Sigma_1^{-1}\mu_1)\right] \\
a &= \boxed{x^T A x + w^T x + b}
\end{aligned}
$$

Question 2.

Expression for joint likelihood and log likelihood.

$$P(x^{(1)},\dots,x^{(m)},y^{(1)},\dots,y^{(m)}) = \frac{\phi(1-\phi)}{((2\pi)^{n/2}|\Sigma|^{1/2})^2}\exp\left\{-\frac{1}{2}[(x-\mu_0)^T\Sigma^{-1}(x-\mu_0)-(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)]\right\}$$

$$P(x^{(1)},\dots,x^{(m)},y^{(1)},\dots,y^{(m)}) = \frac{\phi(1-\phi)}{(2\pi)^n|\Sigma|}\exp\left\{-\frac{1}{2}[(x-\mu_0-x-\mu_1)^T\Sigma^{-1}(x-\mu_0-x-\mu_1)]\right\}$$

$$P(x^{(1)},\dots,x^{(m)},y^{(1)},\dots,y^{(m)}) = \boxed{\frac{\phi(1-\phi)}{(2\pi)^n|\Sigma|}\exp\left\{\frac{1}{2}[(\mu_0+\mu_1)^T\Sigma^{-1}(\mu_0+\mu_1)]\right\}}$$

$$L = \ln P$$

$$L = \boxed{\ln\frac{\phi(1-\phi)}{(2\pi)^n|\Sigma|}+\frac{1}{2}[(\mu_0+\mu_1)^T\Sigma^{-1}(\mu_0+\mu_1)]}$$

MLE and and second derivative for ϕ.

$$\frac{\mathrm{d}L}{\mathrm{d}\phi} = 0$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}\phi}\ln\frac{\phi(1-\phi)}{(2\pi)^n|\Sigma|}$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}\phi}\ln\phi(1-\phi) - \frac{\mathrm{d}}{\mathrm{d}\phi}\ln(2\pi)^n|\Sigma|$$

$$0 = 1-2\phi$$

$$\phi = \boxed{\frac{1}{2}}$$

$$\frac{\mathrm{d}^2L}{\mathrm{d}\phi^2} = \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\ln\phi(1-\phi) - \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\ln(2\pi)^n|\Sigma|$$

$$\frac{\mathrm{d}^2L}{\mathrm{d}\phi^2} = \frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\ln\phi(1-\phi)$$

$$\frac{\mathrm{d}^2L}{\mathrm{d}\phi^2} = \boxed{-2}$$

MLE and and second derivative for μ_0.

$$\frac{\mathrm{d}L}{\mathrm{d}\mu_0} = 0$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}\mu_0}\left[\frac{1}{2}[(\mu_0+\mu_1)^T\Sigma^{-1}(\mu_0+\mu_1)]\right]$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}\mu_0}\mu_0^T\Sigma^{-1}\mu_0 + \frac{\mathrm{d}}{\mathrm{d}\mu_0}\mu_1^T\Sigma^{-1}\mu_1$$

$$0 = \frac{\mathrm{d}}{\mathrm{d}\mu_0}\mu_0^T\Sigma^{-1}\mu_0$$

$$0 = 2\mu_0^T\Sigma^{-1}$$

uhh that didn't work

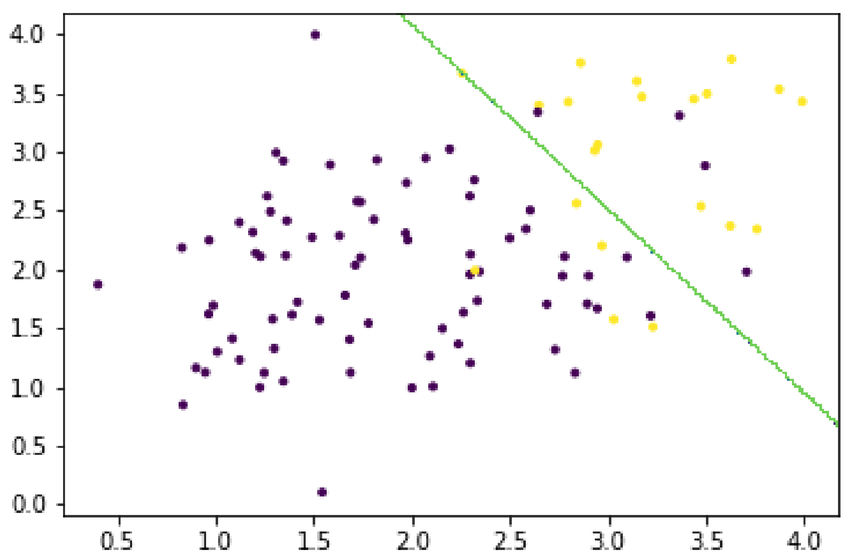
Question 3.

MLEs of parameters for each class.

$$
\begin{aligned}
P(y=0) &= 0.79 \\
\mu_{0,GPA} &= 1.867 \\
\mu_{0,GRE} &= 1.967 \\
\mu_{1,GPA} &= 3.163 \\
\mu_{1,GRE} &= 2.958 \\
\Sigma &= \begin{pmatrix} 0.4456 & 0.0731 \\ 0.0731 & 0.4745 \end{pmatrix}
\end{aligned}
$$

Decision boundary for the linear GDA.

$$
\begin{aligned}
w &= (2.6314, 1.6845) \\
b &= 10.7691
\end{aligned}
$$



Question 4.

Minimum value of the objective function is 0 because the cluster centers would be on top of the points themselves, so all cluster distances to each center would be 0.

Optimal μ_k for regularised k-means and $k = n$. The regularisation will decrease the mean vector's size by bringing it closer to 0, as if there were λ extra points at the origin which were part of each cluster.

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}\mu_n}\left(\lambda|\mu_n|^2 + r_{nk}|x_n - \mu_n|^2\right) \\
0 &= \frac{\mathrm{d}}{\mathrm{d}\mu_n}\left(\lambda\mu_n^T\mu_n + r_{nk}(x_n - \mu_n)^T(x_n - \mu_n)\right) \\
0 &= 2\lambda\mu_n^T - 2r_{nk}(x_n - \mu_n)^T \\
0 &= \lambda\mu_n^T - r_{nk}x_n^T + r_{nk}\mu_n^T \\
\mu_n^T &= \frac{r_{nk}x_n^T}{\lambda + r_{nk}} \\
\mu_k &= \boxed{\frac{\sum_n r_{nk}x_k}{\lambda + \sum_n r_{nk}}}
\end{aligned}$$

Question 5.

skip

Question 6.

A is positive definite for the following value of a.

$$
\begin{aligned}
z^T A z &= \begin{pmatrix} z_1 & z_2 \end{pmatrix} \begin{pmatrix} 9 & 6 \\ 6 & a \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\
0 &\leq \begin{pmatrix} 9z_1 + 6z_2 & 6z_2 + az_1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\
0 &\leq 9z_1^2 + 6z_1 z_2 + 6z_1 z_2 + az_2^2 \\
0 &\leq 12z_1 + az_2 \\
a &\geq \boxed{-12z_1/z_2}
\end{aligned}
$$

Proof that inverse of positive definite matrix is also positive definite. Note that since a positive definite matrix is invertible, it spans R^n, and every vector y can be expressed as $y = \pm Bz$ for a unique z.

$$
\begin{aligned}
0 &\leq z^T B z \\
0 &\leq z^T I B z \\
0 &\leq z^T B^T B^{-1} B z \\
0 &\leq (Bz)^T B^{-1} (Bz) \\
0 &\leq \boxed{y^T B^{-1} y}
\end{aligned}
$$

Data covariance matrix S is positive semi-definite.

$$
\begin{aligned}
z^T S z &= z^T \left(\frac{\Sigma_i (x_i - \mu)(x_i - \mu)^T}{n} \right) z \\
z^T S z &= \frac{1}{n} \Sigma_i z^T (x_i - \mu)(x_i - \mu)^T z \\
z^T S z &= \frac{1}{n} \Sigma_i ((x_i - \mu)^T z)^T ((x_i - \mu)^T z) \\
z^T S z &= \frac{\sum |(x_i - \mu)^T z|^2}{n}
\end{aligned}
$$

Norm of a vector is always at least 0 so $\boxed{z^T S z \geq 0}$.

Question 7.

skip