Students that I collaborated with: Students on ED, Various Students in office hours. Madeline (never got last name / SID: 3036751976

Honor Code.

I certify that all solutions are entirely my own and that I have not looked at anyone else's solution. I have given credit to all external sources I consulted.

X

2) 1)
$$\frac{P(X|Y=C_1)\cancel{P(Y=C_1)}}{\cancel{P(X)}} = \frac{P(X|Y=C_2)\cancel{P(Y=C_2)}}{\cancel{P(X)}}$$

$$N(\mu_i, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$$\cancel{\frac{1}{\sigma\sqrt{2\pi}}} \cdot e^{\frac{-(x-\mu_1)^2}{2\sigma^2}} = \cancel{\frac{1}{\sigma\sqrt{2\pi}}} \cdot e^{\frac{-(x-\mu_2)^2}{2\sigma^2}}$$

$$\frac{\cancel{-}(x-\mu_1)^2}{\cancel{2\sigma^2}} = \frac{\cancel{-}(x-\mu_2)^2}{\cancel{2\sigma^2}}$$

$(x-\mu_1)(x-\mu_1)$

$$(x-\mu_1)^2 = (x-\mu_2)^2$$

$$\cancel{x^2} - 2x\mu_1 + \mu_1^2 = \cancel{x^2} + 2x\mu_2 + \mu_2^x$$

$$-2x\mu_1 - 2x\mu_2 = \mu_2^2 - \mu_1^2$$

$$x(-2\mu_1 - 2\mu_2) = \mu_2^2 - \mu_1^2$$

$$\boxed{x = \frac{\mu_2^2 - \mu_1^2}{-2\mu_1 - 2\mu_2}}$$

2) The probability of a point C_1 being on the wrong side of the boundary means it needs to be to the right of b, so b to ∞. The PDF of this is as shown

$$\int_b^\infty \frac{1}{2\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)dx$$

For C_2 being misclassified, C_2 needs to be on the C_1 side, meaning $-\infty$ to b. So the PDF of that is

$$\int_{-\infty}^{b} \frac{1}{2\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu_2)^2}{2\sigma^2}\right)dx$$

Putting those two together and factoring out an $\frac{1}{2\sqrt{2\pi}\sigma}$ gives us the equation

Shown

3) b^* same as answer in part 1

$$\boxed{b^* = \frac{\mu_2^2 - \mu_1^2}{-2\mu_1 - 2\mu_2}}$$

3) 1) With those conditions, the program will only choose i if it is the most probable choice. The other condition also shows that if the probability is less than a constant, which represents a threshold of the doubt penalty vs. the failure penalty

When $\lambda_r \leq \lambda_s$ then $1 - \lambda_r / \lambda_s$ will be greater than 0.5 so the program will choose doubt more often due to the probability needs to be higher.

2) If $\lambda_r = 0$, that means there is no risk in choosing c+1 or doubt, so the model will be very likely to choose it since it will encurr no loss

Likewise, if $\lambda_r > \lambda_s$ then doubting would be worse off that missclassifing, so the model will choose a classification in order to minimize risk.

4) 1)

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2_i}{2\sigma^2}}$$

ln
do
$\frac{d}{d\mu}$ $\frac{d}{d\sigma}$
set $=0$
and solve
for μ and
σ

$$\sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) + \ln\left(e^{\frac{-(x_i-\mu)^2_i}{2\sigma^2}}\right)$$

$$\sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) + \frac{-(x_i-\mu)^2_i}{2\sigma^2}$$

$$\frac{\partial}{\partial\mu} = \sum_{i=1}^{n} \frac{-\cancel{2}(x_i-\mu)_i}{\cancel{2}\,\sigma^2}$$

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

$$\frac{(x_i-\mu)_i}{\sigma^2} = 0 \qquad \frac{x_i i}{\sigma^2} = \frac{\mu_i}{\sigma^2} \qquad \boxed{\mu = x_i}$$

$$\frac{\partial}{\partial\sigma} = \sum_{i=1}^{n} \frac{-1}{\sqrt{2\pi}}\sigma^2 + \cancel{\frac{-(x_i-x_i)^2_i}{\sigma^3}} = 0$$

$$\sum_{i=1}^{n} \frac{-1}{\sqrt{2\pi}}\sigma^2 = 0$$

$$\boxed{\sigma^2 = 0}$$

2)

$$\int f(x) \cdot p(x)\, dx$$

$\hat{\mu}$ in terms of x

PDF of x

$$\int x_i \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\frac{e^{\frac{\mu}{2\sigma}}}{\sqrt{2}\sqrt{\pi}\,\sigma} \int x\, e^{-\frac{x}{2\sigma}} \qquad u = -\frac{x}{2\sigma} \quad du = -\frac{1}{2\sigma}\, dx$$

$$4\sigma^2 \int u e^u\, du$$

$$u e^u - \int e^u\, du$$

$$-2\sigma x e^{-\frac{x}{2\sigma}} - 4\sigma^2 e^{-\frac{x}{2\sigma}}$$

$$= \frac{e^{\frac{\mu}{2\sigma}}}{\sqrt{2}\sqrt{\pi}\,\sigma} \int x e^{-\frac{x}{2\sigma}}\, dx$$

$$= -\frac{2^{3/2} x e^{\frac{\mu}{2\sigma} - \frac{x}{2\sigma}}}{\sqrt{\pi}} - \frac{2^{\frac{3}{2}} \sigma e^{\frac{\mu}{2\sigma} - \frac{x}{2\sigma}}}{\sqrt{\pi}} = 0$$

unbiased

3)

$$\int f(x) \cdot p(x)\, dx$$

σ^2 in terms of x

PDF of x

$$\int f(x) \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Subbing in this which is σ^2 in terms of x, and Solving this integral.
If integral equals 0, then unbiased, else, biased

4) Risk is expectation of Loss function

$(\hat{\mu} - \mu)^2$

$$E[(\hat{\mu}-\mu)^2] = Var(\hat{\mu})$$

$$Var(\hat{\mu}) = E[\hat{\mu}^2] - E[\overset{L}{\hat{\mu}}]^2$$

$$E[X_i^2] - E[X_i]^2$$

5) 1) X_i is singular, then it does not span across the entire space. This as a result means the columns that make up X_i don't cover the entire space and thus, cannot be linearly independent, and thus means the matrix is not invertible. This means $\hat{\Sigma}_i^{-1}$ will also not be linearly independent and not invertible.

2) We will take all the colums that are not linearly independent and will add on a very small constant. A good way of doing this is to add 0.0001% of the column to each column. This ensures there is minimal error due to adding.

3) The vector x that maximizes the PDF

$$f(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left(-((x-\overset{0}{\cancel{\mu}})^T \Sigma_i^{-1} (x-\overset{0}{\cancel{\mu}})\right)/2$$

$$f(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left(-x^T \Sigma_i^{-1} x / 2\right)$$

The vector that gives the largest value of

$$\frac{-x^T \Sigma_i^{-1} x}{2}$$

would maximize the PDF.

4) $\mathrm{Var}(p) = E[(p - \cancel{E(p)}^{0})(p - \cancel{E(p)}^{0})^T]$

$$E[p p]$$

$$E[y^T x x^T y]$$

$$y^T y$$

$$y^T y \underbrace{E[x x^T]}_{\Sigma_1}$$

$$\boxed{y^T y \Sigma_1}$$

The max eigenvalue will give us the max $y^T X$ or projection on to the unit vector.

8)2) Digit 0. The covariance in the diagonal and off-diagonal terms are larger, with the diagonal terms generally being the largest. This means the variance and covariance of the feature are linked together.

Code on next page

3) c) LDA performed better. My guess would be the digits are easier to handle if you handle the boundaries as straight lines rather than curves.

d) 1 was the easiest digit to classify. This is most likely because of how simple it is compared to other numbers.

Code on next page

4) Kaggle Username: Christopher Avakian
Kaggle Score : 0.845

Code on next page

5) Kaggle Username: Christopher Avakian
Kaggle Score : 0.376 (Yeah IDK how I did that)

Code on next page