1. Given a training dataset $\mathcal{X} = \{x^{(t)}, r^{(t)}\}_{t=1}^{N}$ where $x^{(t)} \in \mathbb{R}$ are scalars and the number of classes is $K = 2$. Suppose instances are normally distributed within each class. Write in closes-form the decision boundary $z \in \mathbb{R}$, where $P[C_1|z] = P[C_2|z]$.

A1: $\because$ $\mathrm{P}(C_1|\mathrm{Z}) = \mathrm{P}(C_2|\mathrm{Z})$ $\because$ Suppose instance are normally distributed within each class

$\therefore$ $\mathrm{P}(\mathrm{Z}|C_1) * P(C1) = \mathrm{P}(\mathrm{Z}|C_2) * \mathrm{P(C2)}$

$$\therefore \frac{1}{\sqrt{2\pi\sigma_1{}^2}} e^{-\frac{(z-u_1)^2}{2\sigma_1{}^2}} * \mathrm{P(C1)} = \frac{1}{\sqrt{2\pi\sigma_2{}^2}} e^{-\frac{(z-u_2)^2}{2\sigma_2{}^2}} * \mathrm{P(C2)}$$

$$\therefore e^{-\frac{(z-u_1)^2}{2\sigma_1{}^2}+\frac{(z-u_2)^2}{2\sigma_2{}^2}} = \frac{\sqrt{2\pi\sigma_1{}^2}}{\sqrt{2\pi\sigma_2{}^2}} * \frac{\mathrm{P(C2)}}{\mathrm{P(C1)}}$$

$\therefore -\frac{(z-u_1)^2}{2\sigma_1{}^2} + \frac{(z-u_2)^2}{2\sigma_2{}^2} = \log\left(\sqrt{2\pi\sigma_1{}^2}\right) - \log\left(\sqrt{2\pi\sigma_2{}^2}\right) + \log(\mathrm{P(C2)}) - \log(\mathrm{P(C1)})$=Q（令 Q 為常數）

$\therefore$boundary 為 一 雙曲線.

$$z = \frac{u_1+u_2}{2} + \frac{1}{2}\left(\sqrt{2\sigma_1{}^2 + log\,\frac{\mathrm{P(C1)}}{\sqrt{2\pi\sigma_1{}^2}}} + \sqrt{2\sigma_2{}^2 + log\,\frac{\mathrm{P(C2)}}{\sqrt{2\pi\sigma_2{}^2}}}\right)$$

2. Consider the univariate parametric classification. Show that the priors $P[C_i]$, $i = 1, 2, \cdots, K$, for different classes can be estimated jointly by assuming that $P[C_i]$ follows a Multinomial distribution parametrized by $\theta = (\rho_1, \cdots, \rho_K)$ with constrains $\sum_{i=1}^{K} \rho_i = 1$ and by maximizing the likelihood $P[\mathcal{X}|\theta]$.

A2:$\because$ $\mathrm{P}(C_i)$follows Multinomial distribution , $\because \sum_{i=1}^{K} \rho_i = 1$

$$\mathrm{P(X)} = \prod_{i=1}^{K} \rho_i{}^{r_i} \ , \ r_i = \begin{cases} 1 & ,X_i\ is\ \ the\ state\ i \\ 0 & ,otherwise \end{cases}$$

$$\log(X|\Theta) = log\left(\prod_{t=1}^{N}\prod_{i=1}^{K} \rho_i{}^{(r_i)^t}\right) = \sum_{t=1}^{N}\sum_{i=1}^{K} \log\left(\rho_i{}^{(r_i)^t}\right) = \sum_{t=1}^{N}\sum_{i=1}^{K} \mathrm{lo}(r_i)^t * \mathrm{g}(\rho_i)$$

Use Lagrange Multiplier -> L = $\sum_{t=1}^{N}\sum_{i=1}^{K} (r_i)^t * \log(\rho_i) + \lambda\left(\sum_{i=1}^{K} \rho_i - 1\right)$

1 --- $\frac{\partial L}{\partial \rho_i} = \sum_{t=1}^{N} (r_i)^t * \left(\frac{1}{\rho_i}\right) + \lambda = 0$, $\sum_{i=1}^{K} (r_i)^t * \log(\rho_i) \rightarrow r_1, r_2 \ldots r_K$ *只有第i個會為* 1 *其他為* 0

$= \left(\frac{1}{\rho_i}\right) * \sum_{t=1}^{N} (r_i)^t + \lambda = 0 \rightarrow \rho_i = -\frac{1}{\lambda}\sum_{t=1}^{N} (r_i)^t$

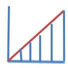
2 --- $\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{K} \rho_i - 1 = 0 \rightarrow \sum_{i=1}^{K} \rho_i = \mathbf{1} = \sum_{i=1}^{K}\left(-\frac{1}{\lambda}\right) * \sum_{t=1}^{N} (r_i)^t = \left(-\frac{1}{\lambda}\right) * \sum_{t=1}^{N}\sum_{i=1}^{K} (r_i)^t$

$= \left(-\frac{1}{\lambda}\right)\sum_{t=1}^{N}(1) = \left(-\frac{1}{\lambda}\right) * N = -\frac{N}{\lambda} \rightarrow \mathbf{N} = -\boldsymbol{\lambda}$

$$\therefore \rho_i = \frac{\sum_{t=1}^{N} (r_i)^t}{N}$$

3. Show that the *Area Under the ROC Curve* (AUC) is equal to the probability that a classifier ranks a randomly chosen positive instance higher than a randomly chosen negative one. (Hint: by partitioning the AUC horizontally)

A3: y(T) is y-value at T at ROC curve , x'(T) is x-value(a point) at T at ROC curve(切無限的直刀,or pdf)



$P_0(T)$ *is* (負類別判斷成正類別的機率,在 T)

AUC=$\int_{-\infty}^{\infty} \mathrm{y(T)} * \mathrm{x'(T)}\ dT$, T *is the threshold* Θ

$= \int_{-\infty}^{\infty} TPR(\mathrm{T}) * \mathrm{FPR'(T)}\ dT = \int_{-\infty}^{\infty} TPR(\mathrm{T}) * P_0(T)\ dT$, (數值*機率=期望值)

$= TPR =$ 若隨機抽取一個正樣本和一個負樣本，分類器依據計算所得到的分數,將這個正樣本排在負樣本前面機率

4. [Coding Assignment] Train a classifier using the probabilistic models such that it is able to predict labels given new examples. You are free to apply any data preprocessing technique, model selection technique, and ensemble method you have learned to improve the prediction accuracy. See coding spec for more details.