Project 0 Theory

Task Description:

we are given a dataset $X = \{x_1, \dots, x_n\}$ which are sampled iid from one of the following distributions.

① Normal Distribution:

$$P_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2} x^2\right), \quad \sigma = \sqrt{2}$$

② Laplace Distribution:

$$P_2(x) = \frac{1}{2b} \exp\left(-1/b\,|x|\right), \quad b = 1$$

③ Student-t distribution:

$$P_3(x) = \frac{\Gamma((v+1)/2)}{\sqrt{v\pi}\,\Gamma(v/2)} \left(1 + \frac{x^2}{v}\right)^{-(v+1)/2}, \quad v = 4$$

35% of the time, the data set is normal,
25% " " " , " " " is Laplace,
40% " " " , " " " is Student t.

let H_i be the event that the data was sampled from P_i, $i = 1, 2, 3$.

Task:

Implement a Bayes-optimal predictor that given the data set X, outputs the probabilities $p(H_i \mid X)$, $i = 1, \dots, n$.

$$+ \quad p(H_i|X) = \frac{p(X|H_i) \,.\, p(H_i)}{\underbrace{p(X)}_{\sum_{i=1}^{3} p(X|H_i) \,.\, p(H_i)}}$$ ← using Baye's rule.

$$\log(p(H_i|X)) = \log(P(X|H_i)) + \log(p(H_i)) - \log\left(\sum_{i=1}^{3} P(X|H_i) \,.\, p(H_i)\right)$$

$$\log(p(H_i|X)) = \log(p(X|H_i)) + \log(p(H_i)) - \log\left(\sum_{i=1}^{3} P(X|H_i) \cdot p(H_i)\right)$$

$i=1:$

$$\log(p(H_1|X)) = \underbrace{\log(p(X|H_1))}_{①} + \underbrace{\log(p(H_1))}_{②} - \log\Big(\underbrace{\sum_{i=1}^{3} P(X|H_i) \,.\, p(H_i)}_{③}\Big)$$ ← can be calculated at the end.

① $P(X|H_1) = p(X_1=x_1, X_2=x_2, \ldots, X_n=x_n | H_1)$

by the iid assumption,

$$= \prod_{j=1}^{n} p(X_j=x_j | H_1) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2} x^2\right)$$

Taking the log:

$$\log(P(X|H_1)) = \sum_{j=1}^{n} \log\left(\frac{1}{2\pi\sigma}\right) - \frac{1}{2\sigma^2} x_j^2$$

$$= n \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} x_j^2$$

$$= n \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2\sigma^2} X^T X$$

② $p(H_1) = 0.25$

$\log(p(H_1)) = \log(0.25)$

$$\therefore \log(p(H_1|X)) = n \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2\sigma^2} \overbrace{X^T X}^{\text{vectorized, efficient}} + \log(.25) - \log(z')$$

where z' is a normalisation factor.

$i = 2$

$$p(X|H_2) = p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | H_2)$$

by the iid assumption,

$$= \prod_{j=1}^{n} p(X_j = x_j | H_2) = \prod_{j=1}^{n} \frac{1}{2b} \exp\left(-\frac{1}{b}|x_j|\right)$$

Taking the log:

$$\log(p(X|H_2)) = \sum_{i=1}^{n} \log\left(\frac{1}{2b}\right) + \log(\exp(-1/b\,|x_j|))$$

$$= n \log\left(\frac{1}{2b}\right) + \sum_{i=1}^{n} -\frac{1}{b}|x_j|$$

$$= n \log\left(\frac{1}{2b}\right) + -\frac{1}{b} \underbrace{\sum_{i=1}^{n} |x_j|}_{l_1\text{-norm}}$$

$$= n \log\left(\frac{1}{2b}\right) - \frac{1}{b} \underbrace{\|X\|_1}_{\text{vectorized, efficient.}}$$

- $p(H_2) = 0.35$

$\log(p(H_2)) = \log(0.35)$

$\therefore\ P(H_2 \mid X) = n\log\left(\frac{1}{2b}\right) - \frac{1}{b}\|X\|_1 + \log(0.35)$
$- \log(z')$

where z' is a normalisation factor.

$i=3$

$P(X \mid H_3) = p(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid H_3)$

by the iid assumption,

$$= \prod_{j=1}^{n} P(X_j = x_j \mid H_3) = \prod_{j=1}^{n} \gamma \left(1 + \frac{x_j^2}{\nu}\right)^{-\frac{(\nu+1)}{2}}, \quad \nu = 4$$

Taking the log:

$$\log(p(X \mid H_3)) = \sum_{j=1}^{n} \log(\gamma) - \frac{(\nu+1)}{2}\log\left(1 + \frac{x_j^2}{\nu}\right)$$

$$= n\log(\gamma) - \left(\frac{\nu+1}{2}\right)\sum_{j=1}^{n}\log\left(1 + \frac{x_j^2}{\nu}\right)$$

where $\gamma = \dfrac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\ \Gamma(\nu/2)}$

- $p(H_3) = 0.4$

$\log(p(H_3)) = \log(0.4)$

$$\therefore \log(p(H_3 \mid X)) = n\log(\gamma) - \left(\frac{v+1}{2}\right)\sum_{j=1}^{n}\log\left(1+\frac{x_j^2}{v}\right) - \log(z')$$

Normalization

$$\log(z') = \overbrace{n\log(\gamma) - \frac{v+1}{2}\sum_{j=1}^{n}\log\left(1+\frac{x_j}{v}\right) + \log(0.4)}^{3}$$

$$\underbrace{+\,n\log\left(\frac{1}{2b}\right) - \frac{1}{2}\|x\|_1 + \log(0.35)}_{2} \; +$$

$$\underbrace{n\log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2\sigma^2}x^T x + \log(.25)}_{1}$$