Statistical Foundations of Learning - Sample Problems 1

CIT4230004 (Summer Semester 2024)

Sample Problem 1.1: Probability Bounds 1

Remind yourself of the Markov inequality to prove the following. If Z is a random variable that takes values in the interval $[0, 1]$, then for $a \in (0, 1)$,

$$\frac{\mathbb{E}[Z] - a}{1 - a} \leq \mathbb{P}(Z > a) \leq \frac{\mathbb{E}[Z]}{a}$$

Proof:
Recall the Markov inequality:

$$\mathbb{P}(Z \geq a) \leq \frac{\mathbb{E}[Z]}{a}$$

For Z taking values in $[0, 1]$, we use the complement probability:

$$\mathbb{P}(Z \leq a) \geq 1 - \mathbb{P}(Z > a)$$

Let $\mathbb{P}(Z > a) = p$. Then,

$$\mathbb{E}[Z] = \int_0^1 z f_Z(z)\, dz$$

$$\mathbb{E}[Z] \geq \int_a^1 z f_Z(z)\, dz \geq a \cdot \mathbb{P}(Z > a) = a \cdot p$$

Hence,

$$\mathbb{P}(Z > a) \leq \frac{\mathbb{E}[Z]}{a}$$

To prove the lower bound, we use:

$$\mathbb{E}[Z] = \int_0^1 z f_Z(z)\, dz$$

$$\mathbb{E}[Z] \leq a \cdot \mathbb{P}(Z \leq a) + \mathbb{P}(Z > a)$$

Thus,

$$\mathbb{E}[Z] \leq a \cdot (1 - p) + p = a + p(1 - a)$$

Rearranging gives:

$$p \geq \frac{\mathbb{E}[Z] - a}{1 - a}$$

So,

$$\frac{\mathbb{E}[Z] - a}{1 - a} \leq \mathbb{P}(Z > a) \leq \frac{\mathbb{E}[Z]}{a}$$

Sample Problem 1.2: Probability Bounds 2

Remind yourself of the Cauchy-Schwarz inequality to prove the following. If Z is a non-negative random variable with finite variance, then

$$\mathbb{E}\left[Z\mathbf{1}\{Z > 0\}\right] \leq \sqrt{\mathbb{E}[Z^2]\mathbb{P}(Z > 0)}$$

Proof:
Recall the Cauchy-Schwarz inequality:

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

Let $X = Z\mathbf{1}\{Z > 0\}$ and $Y = \mathbf{1}\{Z > 0\}$, then:

$$\mathbb{E}[X^2] = \mathbb{E}[Z^2\mathbf{1}\{Z > 0\}] = \mathbb{E}[Z^2]$$

$$\mathbb{E}[Y^2] = \mathbb{E}[\mathbf{1}\{Z > 0\}] = \mathbb{P}(Z > 0)$$

Applying Cauchy-Schwarz:

$$\mathbb{E}\left[Z\mathbf{1}\{Z > 0\}\right] \leq \sqrt{\mathbb{E}[Z^2]\mathbb{P}(Z > 0)}$$

From the above relation, show that:

$$\mathbb{P}(Z > 0) \geq \frac{(\mathbb{E}[Z])^2}{\mathbb{E}[Z^2]}$$

Using the Cauchy-Schwarz result, we have:

$$\mathbb{E}[Z] \leq \sqrt{\mathbb{E}[Z^2]\mathbb{P}(Z > 0)}$$

Squaring both sides:

$$(\mathbb{E}[Z])^2 \leq \mathbb{E}[Z^2]\mathbb{P}(Z > 0)$$

Rearranging gives:

$$\mathbb{P}(Z > 0) \geq \frac{(\mathbb{E}[Z])^2}{\mathbb{E}[Z^2]}$$

Sample Problem 1.3: Probability Bounds 3

We now look at a probability bound defined by the first two moments. Assume that $Z \geq 0$ is a random variable with finite variance and $0 \leq \theta \leq 1$. Show that:

$$\mathbb{P}(Z > \theta \mathbb{E}[Z]) \geq \frac{(1-\theta)^2 (\mathbb{E}[Z])^2}{\mathbb{E}[Z^2]}$$

Proof:
Using Chebyshev's inequality:

$$\mathbb{P}(Z \geq \theta \mathbb{E}[Z]) = \mathbb{P}(Z - \mathbb{E}[Z] \geq (\theta - 1)\mathbb{E}[Z])$$

Define $Y = Z - \mathbb{E}[Z]$:

$$\mathbb{P}(Y \geq (\theta - 1)\mathbb{E}[Z])$$

Using Chebyshev's inequality:

$$\mathbb{P}(Y \geq (\theta - 1)\mathbb{E}[Z]) \geq \frac{(\theta-1)^2 (\mathbb{E}[Z])^2}{\mathbb{E}[Z^2] - (\mathbb{E}[Z])^2}$$

Simplifying gives:

$$\mathbb{P}(Z > \theta \mathbb{E}[Z]) \geq \frac{(1-\theta)^2 (\mathbb{E}[Z])^2}{\mathbb{E}[Z^2]}$$

Sample Problem 1.4: Bayes Risk for 2 Gaussians

Let $X = \mathbb{R}$ and $Y = \{\pm 1\}$. Assume $Y \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, and

$$X \sim N(\mu, \sigma^2), \text{ if } Y = 1$$

$$X \sim N(-\mu, \sigma^2), \text{ if } Y = -1$$

for some $\mu, \sigma^2 > 0$. Derive $\eta(x)$, the Bayes classifier h^*, and the Bayes risk. Express everything in terms of the probability density function ϕ and the cumulative distribution function Φ of Gaussian random variables.

Solution:
The posterior probability:

$$\eta(x) = \mathbb{P}(Y = 1 | X = x)$$

Using Bayes' theorem:

$$\eta(x) = \frac{\mathbb{P}(X = x | Y = 1)\mathbb{P}(Y = 1)}{\mathbb{P}(X = x)}$$

The likelihoods:

$$\mathbb{P}(X = x | Y = 1) = \phi(x; \mu, \sigma^2)$$

$$\mathbb{P}(X = x|Y = -1) = \phi(x; -\mu, \sigma^2)$$

The priors:

$$\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = \frac{1}{2}$$

The marginal:

$$\mathbb{P}(X = x) = \frac{1}{2}\phi(x; \mu, \sigma^2) + \frac{1}{2}\phi(x; -\mu, \sigma^2)$$

Thus:

$$\eta(x) = \frac{\phi(x; \mu, \sigma^2)}{\phi(x; \mu, \sigma^2) + \phi(x; -\mu, \sigma^2)}$$

The Bayes classifier:

$$h^*(x) = \begin{cases} 1 & \text{if } \eta(x) > \frac{1}{2} \\ -1 & \text{otherwise} \end{cases}$$

Using the expression for $\eta(x)$:

$$\eta(x) > \frac{1}{2} \implies \frac{\phi(x; \mu, \sigma^2)}{\phi(x; \mu, \sigma^2) + \phi(x; -\mu, \sigma^2)} > \frac{1}{2}$$

Simplifying:

$$\phi(x; \mu, \sigma^2) > \phi(x; -\mu, \sigma^2)$$

Since the Gaussian density function is symmetric:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} > \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x+\mu)^2}{2\sigma^2}}$$

$$e^{-\frac{(x-\mu)^2}{2\sigma^2}} > e^{-\frac{(x+\mu)^2}{2\sigma^2}}$$

Taking the natural logarithm on both sides:

$$-\frac{(x-\mu)^2}{2\sigma^2} > -\frac{(x+\mu)^2}{2\sigma^2}$$

$$(x-\mu)^2 < (x+\mu)^2$$

Expanding and simplifying:

$$x^2 - 2x\mu + \mu^2 < x^2 + 2x\mu + \mu^2$$

$$-2x\mu < 2x\mu$$

$$x < 0$$

Therefore, the Bayes classifier is:

$$h^*(x) = \begin{cases} 1 & \text{if } x < 0 \\ -1 & \text{if } x \geq 0 \end{cases}$$

To find the Bayes risk, we compute the probability of error:

$$R(h^*) = \mathbb{P}(h^*(X) \neq Y)$$

This is the probability that $h^*(X)$ makes an incorrect prediction:

$$R(h^*) = \mathbb{P}(X \geq 0 \mid Y = 1)\mathbb{P}(Y = 1) + \mathbb{P}(X < 0 \mid Y = -1)\mathbb{P}(Y = -1)$$

Given $X \sim N(\mu, \sigma^2)$ if $Y = 1$ and $X \sim N(-\mu, \sigma^2)$ if $Y = -1$:

$$\mathbb{P}(X \geq 0 \mid Y = 1) = 1 - \Phi\left(\frac{-\mu}{\sigma}\right) = \Phi\left(\frac{\mu}{\sigma}\right)$$

$$\mathbb{P}(X < 0 \mid Y = -1) = \Phi\left(\frac{\mu}{\sigma}\right)$$

Since $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = \frac{1}{2}$, we have:

$$R(h^*) = \Phi\left(\frac{\mu}{\sigma}\right) \cdot \frac{1}{2} + \Phi\left(\frac{\mu}{\sigma}\right) \cdot \frac{1}{2}$$

$$R(h^*) = \Phi\left(\frac{\mu}{\sigma}\right)$$

Thus, the Bayes risk is:

$$R(h^*) = \Phi\left(\frac{\mu}{\sigma}\right)$$