DA 5001/6400 (July-Nov 2024): HW0

Arjav Singh MM20B007
"I ACCEPT THE HONOUR CODE"

August 9, 2024

Problem 1: Gaussian Log-Likelihood Ratio

Given:

- $p(x)$ is the PDF of a Gaussian distribution $N(0, \sigma^2)$.
- $q(x)$ is the PDF of a Gaussian distribution $N(\mu, \sigma^2)$.

The general form of a Gaussian distribution $N(\mu, \sigma^2)$ is:

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

So, for $p(x)$ and $q(x)$:

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

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

We need to find the PDF of the random variable $Y = \log\left(\frac{p(X)}{q(X)}\right)$.
First, compute the ratio:

$$\frac{p(x)}{q(x)} = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} = \exp\left(-\frac{x^2}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^2}\right)$$

Simplifying the exponent:

$$\frac{p(x)}{q(x)} = \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x^2 - 2\mu x + \mu^2}{2\sigma^2}\right) = \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x^2}{2\sigma^2} - \frac{\mu x}{\sigma^2} + \frac{\mu^2}{2\sigma^2}\right)$$

$$= \exp\left(\frac{-\mu x + \frac{\mu^2}{2}}{\sigma^2}\right) = \exp\left(\frac{-\mu x}{\sigma^2} + \frac{\mu^2}{2\sigma^2}\right)$$

Now, taking the logarithm:

$$Y = \log\left(\frac{p(x)}{q(x)}\right) = -\frac{\mu x}{\sigma^2} + \frac{\mu^2}{2\sigma^2}$$

Since $X \sim N(0, \sigma^2)$, therefore, $-\frac{\mu X}{\sigma^2}$ is normally distributed with mean 0 and variance $\frac{\mu^2}{\sigma^2}$. Let's denote this by $N\left(0, \frac{\mu^2}{\sigma^2}\right)$. Adding the constant $\frac{\mu^2}{2\sigma^2}$ shifts the mean but does not affect the variance. Therefore, the random variable $Y = \log\left(\frac{p(X)}{q(X)}\right)$ is normally distributed with mean $\frac{\mu^2}{2\sigma^2}$ and variance $\frac{\mu^2}{\sigma^2}$. Thus:

$$Y \sim N\left(\frac{\mu^2}{2\sigma^2}, \frac{\mu^2}{\sigma^2}\right)$$

Problem 2: Markov and Chebyshev inequalities

Part 1

Given a non-negative random variable X.
To prove:

$$P(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

.
Let $f(x)$ be the Probability Density Function of X. Then, we can write the expectation of the random variable X can be expressed as:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_{0}^{\infty} x \cdot f(x)\, dx$$

Since, X is non-negative.
For any $t > 0$,

$$\mathbb{E}[X] \geq \int_{t}^{\infty} x \cdot f(x)\, dx$$

Also,

$$\int_{t}^{\infty} x \cdot f(x)\, dx \geq \int_{t}^{\infty} t \cdot f(x)\, dx$$

So,

$$\mathbb{E}[X] \geq t \cdot \int_{t}^{\infty} f(x)\, dx$$

$$\mathbb{E}[X] \geq t \cdot P(X > t)$$

Hence proved,

$$P(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

Part 2

To prove that $P(|X - \mathbb{E}[X]| > t) \leq \frac{\mathrm{Var}(X)}{t^2}$ for a random variable X with finite variance.
Let $Y = (X - \mathbb{E}[X])^2$ be a non-negative random variable. Now, as per Markov's inequality, for any positive number t^2

$$P(Y > t^2) \leq \frac{\mathbb{E}[Y]}{t^2}$$

$$P((X - \mathbb{E}[X])^2 > t^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\mathbb{V}\mathrm{ar}[X]}{t^2}$$

Hence proved the Chebyshev's inequality.

$$P(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathbb{V}\text{ar}[X]}{t^2}$$

Problem 3: Moment Generating Functions

Part 1

We are given a random variable X having a Gaussian distribution $N(0, \sigma^2)$. The moment generating function of a random variable is given by

$$\phi_X(\lambda) = \mathbb{E}\left[e^{\lambda X}\right]$$

The general form of a Gaussian distribution $N(\mu, \sigma^2)$ is:

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

For $\phi_X(\lambda)$,

$$\phi_X(\lambda) = \int_{-\infty}^{\infty} e^{\lambda x} \cdot f(x)\, dx$$

Since $\mu = 0$

$$\phi_X(\lambda) = \int_{-\infty}^{\infty} \exp(\lambda x) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx$$

$$\phi_X(\lambda) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\lambda x - \frac{x^2}{2\sigma^2}\right) dx$$

$$\phi_X(\lambda) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{\lambda^2\sigma^2}{2} - \frac{(x-\lambda\sigma^2)^2}{2\sigma^2}\right) dx$$

$$\phi_X(\lambda) = \exp\left(\frac{\lambda^2\sigma^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\lambda\sigma^2)^2}{2\sigma^2}\right) dx$$

Here, the integral part is a Gaussian Distribution with $\mu = \lambda\sigma^2$ and $Var = \sigma^2$, and integration of it over $(-\infty, \infty) = 1$.
So, the MGF of Gaussian distribution $N(0, \sigma^2)$ is:

$$\phi_X(\lambda) = e^{\frac{\lambda^2\sigma^2}{2}}$$

Part 2

Given a random variable X having a chi-squared distribution with 1 degree of freedom. That is $X \sim N(0, 1)$.
We have to find $\mathbb{E}[e^{\lambda X^2}]$.

$$\mathbb{E}[e^{\lambda X^2}] = \int_{-\infty}^{\infty} e^{\lambda x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

$$\mathbb{E}[e^{\lambda X^2}] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(\lambda - \frac{1}{2})x^2} \, dx$$

The integral part can be seen as a Gaussian Distribution with mean = 0 and variance = $\frac{1}{1-2\lambda}$

$$\mathbb{E}[e^{\lambda X^2}] = \frac{1}{\sqrt{2(\frac{1}{2} - \lambda)}} = \frac{1}{\sqrt{1 - 2\lambda}}$$

So,

$$\mathbb{E}[e^{\lambda X^2}] = \frac{1}{\sqrt{1 - 2\lambda}} for \lambda < \frac{1}{2}$$

.

Problem 4: Convexity of the KL Divergence

Part 1

To show that the function $h : (0, 1) \to \mathbb{R}$ given by $h(x) = x \log(x)$ is convex, we use the condition that a twice differentiable function h is convex if and only if its second derivative $h''(x)$ is non-negative for all x in its domain.
The function is $h(x) = x \log(x)$.
For $h(x) = x \log(x)$, the first derivative is:

$$h'(x) = \frac{d}{dx} (x \log(x)) = \log(x) + x \cdot \frac{1}{x} = \log(x) + 1$$

Next, find the second derivative $h''(x)$ by differentiating $h'(x)$:

$$h''(x) = \frac{d}{dx} (\log(x) + 1) = \frac{1}{x}$$

For convexity, $h''(x)$ must be non-negative for all x in the domain of h. Here, $h''(x) = \frac{1}{x}$.
In the domain $(0, 1)$, $\frac{1}{x}$ is positive because $x > 0$. Therefore, $h''(x) \geq 0$ for all $x \in (0, 1)$.
Since the second derivative $h''(x) = \frac{1}{x}$ is positive for all $x \in (0, 1)$, the function $h(x) = x \log(x)$ is convex on $(0, 1)$.

Part 2

To prove the log-sum inequality using the convexity of $h(x) = x \log(x)$, we start with the function h and use the fact that it is convex.
For $a_1, a_2, b_1, b_2 > 0$, we want to show:

$$(a_1 + a_2) \log \left(\frac{a_1 + a_2}{b_1 + b_2} \right) \leq a_1 \log \left(\frac{a_1}{b_1} \right) + a_2 \log \left(\frac{a_2}{b_2} \right).$$

Rewrite the right-hand side using $h(x) = x \log(x)$:

$$a_1 \log \left(\frac{a_1}{b_1} \right) + a_2 \log \left(\frac{a_2}{b_2} \right) = b_1 \cdot h \left(\frac{a_1}{b_1} \right) + b_2 \cdot h \left(\frac{a_2}{b_2} \right).$$

Let $\lambda = \frac{b_1}{b_1+b_2}$ and $1-\lambda = \frac{b_2}{b_1+b_2}$. Using the convexity of h:

$$h\left(\frac{a_1+a_2}{b_1+b_2}\right) \leq (b_1+b_2)\left(\lambda h\left(\frac{a_1}{b_1}\right) + (1-\lambda) h\left(\frac{a_2}{b_2}\right)\right).$$

Multiplying both sides by b_1+b_2:

$$(a_1+a_2)\log\left(\frac{a_1+a_2}{b_1+b_2}\right) \leq b_1 \cdot h\left(\frac{a_1}{b_1}\right) + b_2 \cdot h\left(\frac{a_2}{b_2}\right).$$

Thus,

$$(a_1+a_2)\log\left(\frac{a_1+a_2}{b_1+b_2}\right) \leq a_1 \log\left(\frac{a_1}{b_1}\right) + a_2 \log\left(\frac{a_2}{b_2}\right).$$

Part 3

To prove that the KL divergence function $D_{KL}(p \parallel q)$ is convex, we need to show that for distributions p, p', q, q', and $\lambda \in [0,1]$, the following inequality holds:

$$D_{KL}(\lambda p + (1-\lambda)p' \parallel \lambda q + (1-\lambda)q') \leq \lambda D_{KL}(p \parallel q) + (1-\lambda) D_{KL}(p' \parallel q')$$

KL divergence between discrete distributions p and q is defined as

$$D_{KL}(p \parallel q) = \sum_{i=1}^{k} p_i \log \frac{p_i}{q_i}$$

$$D_{KL}(\lambda p + (1-\lambda)p' \parallel \lambda q + (1-\lambda)q') = \sum_{i=1}^{k} (\lambda p_i + (1-\lambda)p'_i) \log\left(\frac{\lambda p_i + (1-\lambda)p'_i}{\lambda q_i + (1-\lambda)q'_i}\right)$$

Applying the log-sum inequality:

$$\log\left(\frac{\lambda p_i + (1-\lambda)p'_i}{\lambda q_i + (1-\lambda)q'_i}\right) \leq \frac{\lambda p_i}{\lambda p_i + (1-\lambda)p'_i} \log\left(\frac{p_i}{q_i}\right) + \frac{(1-\lambda)p'_i}{\lambda p_i + (1-\lambda)p'_i} \log\left(\frac{p'_i}{q'_i}\right)$$

$$(\lambda p_i + (1-\lambda)p'_i) \log\left(\frac{\lambda p_i + (1-\lambda)p'_i}{\lambda q_i + (1-\lambda)q'_i}\right) \leq (\lambda p_i) \log\left(\frac{p_i}{q_i}\right) + ((1-\lambda)p'_i) \log\left(\frac{p'_i}{q'_i}\right)$$

Rewriting KL divergence between discrete distributions p and q as:

$$D_{KL}(\lambda p + (1-\lambda)p' \parallel \lambda q + (1-\lambda)q') \leq \sum_{i=1}^{k} \left(\lambda p_i \log\left(\frac{p_i}{q_i}\right) + (1-\lambda)p'_i \log\left(\frac{p'_i}{q'_i}\right)\right)$$

$$D_{KL}(\lambda p + (1-\lambda)p' \parallel \lambda q + (1-\lambda)q') \leq \lambda \sum_{i=1}^{k} p_i \log\left(\frac{p_i}{q_i}\right) + (1-\lambda) \sum_{i=1}^{k} p'_i \log\left(\frac{p'_i}{q'_i}\right)$$

$$D_{KL}(\lambda p + (1-\lambda)p' \parallel \lambda q + (1-\lambda)q') \leq \lambda D_{KL}(p \parallel q) + (1-\lambda) D_{KL}(p' \parallel q')$$

Hence proved.

Problem 5: KL divergence between multivariate Gaussians

Let $P = \mathcal{N}(\mu_1, \Sigma)$ and $Q = \mathcal{N}(\mu_2, \Sigma)$ be two multivariate Gaussian distributions with means $\mu_1, \mu_2 \in \mathbb{R}^d$ and equal covariance $\Sigma \in \mathbb{S}_d^{++}$.
The KL divergence $D_{KL}(P \parallel Q)$ is given by:

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx$$

where the probability density functions are:

$$p(x) = \frac{1}{(2\pi)^{d/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^\top \Sigma^{-1} (x-\mu_1) \right)$$

$$q(x) = \frac{1}{(2\pi)^{d/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_2)^\top \Sigma^{-1} (x-\mu_2) \right)$$

We find:

$$\frac{p(x)}{q(x)} = \exp\left((\mu_1 - \mu_2)^\top \Sigma^{-1} x - \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1) \right)$$

Taking the logarithm:

$$\log \frac{p(x)}{q(x)} = (\mu_1 - \mu_2)^\top \Sigma^{-1} x - \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1)$$

Taking the expectation with respect to $p(x)$:

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \left((\mu_1 - \mu_2)^\top \Sigma^{-1} x - \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1) \right) dx$$

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{\infty} (\mu_1 - \mu_2)^\top \Sigma^{-1} x p(x) \, dx - \int_{-\infty}^{\infty} \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1) p(x) \, dx$$

$$D_{KL}(P \parallel Q) = (\mu_1 - \mu_2)^\top \Sigma^{-1} \int_{-\infty}^{\infty} x p(x) \, dx - \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1) \int_{-\infty}^{\infty} p(x) \, dx$$

$$D_{KL}(P \parallel Q) = (\mu_1 - \mu_2)^\top \Sigma^{-1} \mu_1 - \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1)$$

$$D_{KL}(P \parallel Q) = \frac{1}{2}(\mu_2 - \mu_1)^\top \Sigma^{-1} (\mu_2 - \mu_1)$$

Problem 6: Gradient Descent on Quadratic Functions

Given the quadratic function $f : \mathbb{R}^d \to \mathbb{R}$ given by

$$f(\theta) = \frac{1}{2}\theta^\top A \theta + b^\top \theta,$$

where $A \in \mathbb{S}_{++}^d$ (positive definite matrix) and $b \in \mathbb{R}^d$.

Part 1

To find the gradient of $f(\theta)$:

$$f(\theta) = \frac{1}{2}\theta^\top A\theta + b^\top\theta$$

$$\nabla f(\theta) = \nabla_\theta(\frac{1}{2}\theta^\top A\theta + b^\top\theta) = A\theta + b$$

Part 2

$$\nabla^2 f(\theta) = \nabla_\theta(A\theta + b) = A$$

Since A is positive definite, the function $f(\theta)$ is strictly convex. The global minimizer $\theta^\star$ is given by:

$$\theta^\star = -A^{-1}b.$$

Part 3

Gradient descent with learning rate γ produces the sequence:

$$\theta_{t+1} = \theta_t - \gamma(A\theta_t + b).$$

Rewriting this, we get:

$$\theta_{t+1} = (I - \gamma A)\theta_t - \gamma b = (I - \gamma A)\theta_t - \gamma A\theta^\star.$$

Thus,

$$\theta_{t+1} - \theta^\star = (I - \gamma A)(\theta_t - \theta^\star).$$

So,

$$\theta_t - \theta^\star = (I - \gamma A)^t(\theta_0 - \theta^\star).$$

Part 4

Let λ_1 be the largest eigenvalue of A. The matrix $I - \gamma A$ has eigenvalues $1 - \gamma\lambda_i$ where λ_i are the eigenvalues of A. The largest eigenvalue is λ_1, so:

$$\text{Eigenvalues of } (I - \gamma A) \text{ are } 1 - \gamma\lambda_i \text{ for } i = 1, \ldots, d$$

If $\gamma \leq \frac{1}{\lambda_1}$, then:

$$1 - \gamma\lambda_1 \geq 0$$

In particular, $1 - \gamma\lambda_i$ for all λ_i will be non-negative and less than or equal to 1. Thus, the spectral norm (largest absolute eigenvalue) of $I - \gamma A$ is at most $1 - \gamma\lambda_1$. Since $\gamma \leq \frac{1}{\lambda_1}$, we have:

$$1 - \gamma\lambda_1 \geq 0$$

Thus, $\|(I - \gamma A)^t\| \leq (1 - \gamma\lambda_1)^t$, which tends to 0 as $t \to \infty$. Therefore:

$$\lim_{t\to\infty} \|\theta_t - \theta^\star\|^2 = 0$$

Part 5

If $\gamma > \frac{2}{\lambda_1}$, then:

$$1 - \gamma\lambda_1 < -1$$

For some eigenvalues, $1-\gamma\lambda_i$ will be negative, which means that $\|(I-\gamma A)^t(\theta_0-\theta^\star)\|$ will grow exponentially. More formally:

$$\text{If } 1 - \gamma\lambda_i < -1 \text{ for some } \lambda_i \text{ (i.e., } \gamma > \frac{2}{\lambda_i}\text{)},$$

then $(I - \gamma A)^t$ can result in increasing norms over iterations. Therefore:

$$\lim_{t\to\infty} \|\theta_t - \theta^\star\|^2 = \infty$$

Problem 7: Membership Inference: Toy Problem

Part 1

For dataset D_0, we have:

$$f(D_0) = \sum_{i=1}^{n} x_i$$

$$z^\top f(D_0) = z^\top \left(\sum_{i=1}^{n} x_i\right)$$

Since $x_i \sim \mathcal{N}(0, I_d)$, $\mathbb{E}[x_i] = 0$ for all i, so:

$$\mathbb{E}[z^\top f(D_0)] = \mathbb{E}\left[z^\top \left(\sum_{i=1}^{n} x_i\right)\right] = \sum_{i=1}^{n} \mathbb{E}[z^\top x_i] = 0$$

For dataset D_1, where $D_1 = D_0 \cup \{z\}$, we have:

$$f(D_1) = \sum_{i=1}^{n} x_i + z$$

$$z^\top f(D_1) = z^\top \left(\sum_{i=1}^{n} x_i + z\right) = z^\top \sum_{i=1}^{n} x_i + z^\top z$$

Thus:

$$\mathbb{E}[z^\top f(D_1)] = d\sigma^2$$

Part 2

For D_0, we have:

$$\text{Var}(z^\top f(D_0)) = nd\sigma^2$$
$$\mathbb{E}[(z^\top f(D_0))^2] = nd\sigma^2$$

For D_1, we have:

$$\text{Var}(z^\top f(D_1)) = nd\sigma^2 + 2d\sigma^4$$
$$\mathbb{E}[(z^\top f(D_1))^2] = nd\sigma^2 + 2d\sigma^4 + d^2\sigma^4$$

Part 3

Given $z^\top f(D_0) \sim \mathcal{N}(0, nd\sigma^2)$ and $z^\top f(D_1) \sim \mathcal{N}(d\sigma^2, nd\sigma^2 + 2d\sigma^4)$:
1. Type-I Error:

$$P_{\text{Type-I}} = 1 - \Phi\left(\frac{t}{\sqrt{nd\sigma^2}}\right)$$

2. Type-II Error:

$$P_{\text{Type-II}} = \Phi\left(\frac{t - d\sigma^2}{\sqrt{nd\sigma^2 + 2d\sigma^4}}\right)$$

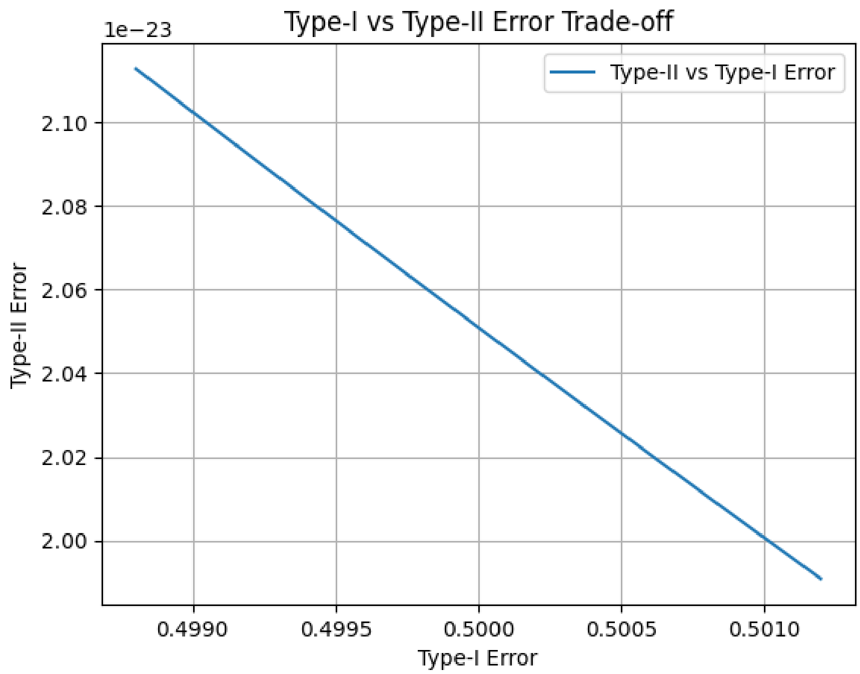

Figure 1: Trade-off between Type-I and Type-II Errors

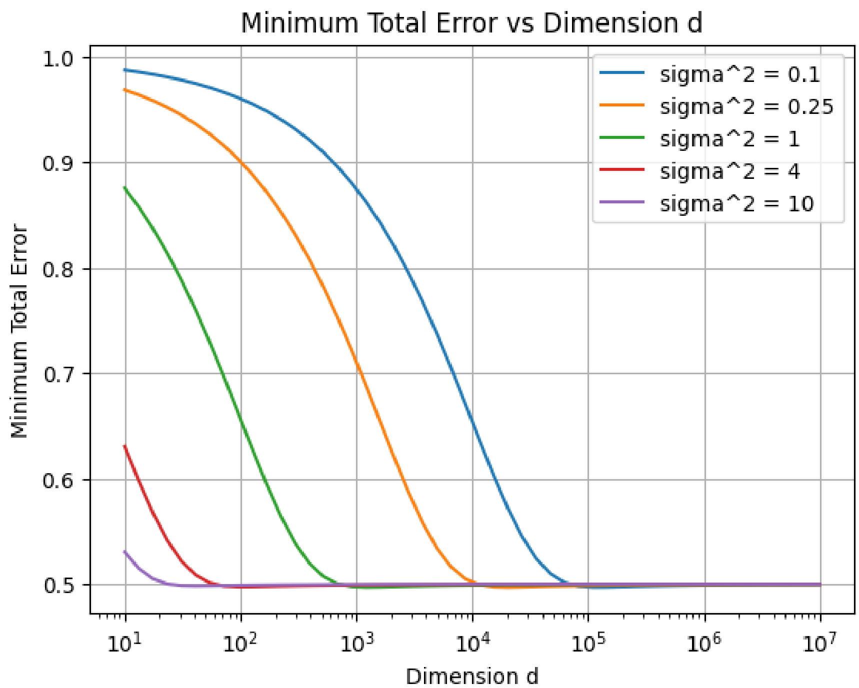

Figure 2: Minimum Total Error vs. Dimension d for different σ^2 values

Part 4

To compute the total error:

$$\text{Total Error} = \mathrm{P}_{\text{Type-I}} + \mathrm{P}_{\text{Type-II}}$$

Part 5

1. Type-I Error: Given that $z \sim N(0, \sigma^2 I_d)$ and $f(D_0) = \sum_{i=1}^{n} x_i$, where $x_i \sim N(0, I_d)$, the expression $z^\top f(D_0)$ is distributed as a Gaussian with mean 0 and variance $n\sigma^2 d$. Therefore:

$$z^\top f(D_0) \sim N\left(0, n\sigma^2 d\right)$$

As d increases, the variance $n\sigma^2 d$ increases, and the Type-I error, which is $P(z^\top f(D_0) > t)$, approaches 1 for any fixed threshold t. This is because the tail probability of a normal distribution with increasing variance tends to be 1.

2. Type-II Error: Similarly, for $z^\top f(D_1)$, we have:

$$z^\top f(D_1) \sim N\left(d\sigma^2, n\sigma^2 d + 2\sigma^4 d\right)$$

The variance here is also increasing with d. The Type-II error, $P(z^\top f(D_1) < t)$, approaches 1 for any fixed t. This is because, even though the mean shifts by $d\sigma^2$, the increasing variance dominates the mean shift as d becomes very large, making the overlap between the distributions for D_0 and D_1 substantial.

As $d \to \infty$, both Type-I and Type-II errors approach 1. In high-dimensional settings, the adversary's task becomes increasingly difficult, leading to a situation where the adversary cannot reliably distinguish between the two hypotheses H_0 and H_1. Essentially, the data becomes too noisy for the adversary to make accurate inferences, highlighting the curse of dimensionality in this context.

Extra Credit 1: Maximum of the Uniform Distribution

Let $X_1, \ldots, X_n \sim \text{Unif}(0, a)$ be n i.i.d. samples distributed between 0 and $a > 0$. Let $Y = \max\{X_1, \ldots, X_n\}$ denote the largest sample.

Part 1

The cumulative distribution function (CDF) of Y is:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \left(\frac{y}{a}\right)^n, \quad 0 \leq y \leq a$$

The probability density function (PDF) of Y is:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{n}{a^n}y^{n-1}, \quad 0 \leq y \leq a$$

The mean of Y is:

$$\mathbb{E}[Y] = \int_0^a y f_Y(y)\, dy = \int_0^a y \frac{n}{a^n} y^{n-1}\, dy = \frac{n}{a^n}\int_0^a y^n\, dy$$

$$\mathbb{E}[Y] = \frac{n}{a^n} \cdot \frac{a^{n+1}}{n+1} = \frac{n}{n+1} \cdot a$$

Thus, the mean of Y is:

$$\mathbb{E}[Y] = \frac{n}{n+1} \cdot a$$

Part 2

To compute the variance, $\text{Var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$, we first find $\mathbb{E}[Y^2]$:

$$\mathbb{E}[Y^2] = \int_0^a y^2 f_Y(y)\, dy = \int_0^a y^2 \frac{n}{a^n} y^{n-1}\, dy = \frac{n}{a^n}\int_0^a y^{n+1}\, dy$$

$$\mathbb{E}[Y^2] = \frac{n}{a^n} \cdot \frac{a^{n+2}}{n+2} = \frac{n}{(n+2)(n+1)} \cdot a^2$$

The variance is then:

$$\text{Var}(Y) = a^2 \left(\frac{n}{(n+2)(n+1)} - \frac{n^2}{(n+1)^2}\right)$$

Part 3

The squared error is given by:

$$\mathbb{E}[(Y-a)^2] = \text{Var}(Y) + (\mathbb{E}[Y]-a)^2$$

Substituting the values:

$$\mathbb{E}[(Y-a)^2] = a^2\left(\frac{n}{(n+2)(n+1)} - \frac{n^2}{(n+1)^2} + \frac{1}{(n+1)^2}\right)$$