

Problem Set 10, Nov 22, 2024 (Solutions to Theory Questions)

1 Vector Calculus

1. We have $\nabla f(\mathbf{x}) = (\mathbf{A} + \mathbf{A}^\top)\mathbf{x} + \mathbf{b}$. One way to see this is to explicitly expand out the expression. We have

$$f(\mathbf{x}) = \sum_{i,j} A_{i,j} x_i x_j + \sum_i b_i x_i + c.$$

If we now take the derivative with respect to x_k we get

$$\frac{\partial f(\mathbf{x})}{\partial x_k} = \sum_j A_{k,j} x_j + \sum_i A_{i,k} x_i + b_k.$$

2. $\nabla^2 f(\mathbf{x}) = \mathbf{A} + \mathbf{A}^\top$. Taking the derivative of $\frac{\partial f(\mathbf{x})}{\partial x_k}$, as given in the previous expression, with respect to x_l we get

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_k \partial x_l} = A_{k,l} + A_{l,k}.$$

2 Maximum Likelihood Principle

1. The likelihood is given by

$$\begin{aligned}\mathbb{P}[X_1, \dots, X_N | \mu, \sigma^2] &= \prod_{n=1}^N \mathbb{P}[X_n | \mu, \sigma^2] \\ &= \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_n - \mu)^2}{2\sigma^2}\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{\sum_{n=1}^N (X_n - \mu)^2}{2\sigma^2}\right)\end{aligned}$$

2. It might be easier to work with the negative log-likelihood, given by

$$\begin{aligned}-\log \mathbb{P}[X_1, \dots, X_N | \mu, \sigma^2] &= -\log \left[\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{\sum_{n=1}^N (X_n - \mu)^2}{2\sigma^2}\right) \right] \\ &= \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{n=1}^N (X_n - \mu)^2 \\ &= \frac{N}{2} \log(2\pi) + \frac{N}{2} \log(\sigma^2) + \frac{1}{2\sigma^2} \sum_{n=1}^N (X_n - \mu)^2\end{aligned}$$

The derivative with respect to μ is

$$\begin{aligned}-\frac{\partial \log \mathbb{P}[X_1, \dots, X_N | \mu, \sigma^2]}{\partial \mu} &= \frac{1}{2\sigma^2} \frac{\partial \left(\sum_{n=1}^N (X_n^2 - 2X_n\mu + \mu^2) \right)}{\partial \mu} \\ &= \frac{1}{2\sigma^2} \sum_{n=1}^N (-2X_n + 2\mu) \\ &= \frac{1}{\sigma^2} \sum_{n=1}^N (-X_n + \mu)\end{aligned}$$

Setting this expression to 0, we get $\hat{\mu} = \frac{1}{N} \sum_{n=1}^N X_n$.

The derivative with respect to σ^2 is

$$\begin{aligned}-\frac{\partial \log \mathbb{P}[X_1, \dots, X_N | \mu, \sigma^2]}{\partial \sigma^2} &= \frac{N}{2} \frac{\partial \log(\sigma^2)}{\partial \sigma^2} + \frac{\partial \frac{1}{\sigma^2}}{\partial \sigma^2} \frac{1}{2} \sum_{n=1}^N (X_n - \mu)^2 \\ &= \frac{N}{2} \frac{1}{\sigma^2} - \frac{1}{\sigma^4} \frac{1}{2} \sum_{n=1}^N (X_n - \mu)^2\end{aligned}$$

Setting this expression to 0, and replacing the unknown quantity μ by the estimate $\hat{\mu}$ we get

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (X_n - \hat{\mu})^2.$$

3. By linearity of expectation, we get $\mathbb{E}[\hat{\mu}] = \frac{1}{N} \sum_{n=1}^N \mathbb{E}[X_n] = \mu$. So indeed, this estimate is *unbiased*.

4. We get that

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \mathbb{E} \left[\frac{1}{N} \sum_{n=1}^N (X_n - \hat{\mu})^2 \right] \\
&= \mathbb{E} \left[\frac{1}{N} \sum_{n=1}^N ((X_n - \mu) - (\hat{\mu} - \mu))^2 \right] \\
&= \mathbb{E} \left[\frac{1}{N} \sum_{n=1}^N \left((X_n - \mu) - \frac{1}{N} \sum_{j=1}^N (X_j - \mu) \right)^2 \right] \\
&= \mathbb{E} \left[\frac{1}{N} \sum_{n=1}^N \left(\frac{N-1}{N} (X_n - \mu) - \frac{1}{N} \sum_{j \neq n} (X_j - \mu) \right)^2 \right] \\
&= \frac{1}{N} \sum_{n=1}^N \mathbb{E} \left[\left(\frac{N-1}{N} (X_n - \mu) - \frac{1}{N} \sum_{j \neq n} (X_j - \mu) \right)^2 \right].
\end{aligned}$$

Since the variables $X_i - \mu$ and $X_j - \mu$ for $i \neq j$ are independent and have mean $= 0$, we can separate out the expectations as

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \frac{1}{N} \sum_{n=1}^N \mathbb{E} \left[\left(\frac{N-1}{N} (X_n - \mu) - \frac{1}{N} \sum_{j \neq n} (X_j - \mu) \right)^2 \right] \\
&= \frac{1}{N} \sum_{n=1}^N \mathbb{E} \left[\left(\frac{N-1}{N} (X_n - \mu) \right)^2 \right] + \frac{1}{N} \sum_{n=1}^N \sum_{j \neq n} \mathbb{E} \left[\left(\frac{1}{N} (X_j - \mu) \right)^2 \right] \\
&= \frac{(N-1)^2}{N^3} \sum_{n=1}^N \mathbb{E} \left[(X_n - \mu)^2 \right] + \frac{1}{N^3} \sum_{n=1}^N \sum_{j \neq n} \mathbb{E} \left[(X_j - \mu)^2 \right] \\
&= \frac{(N-1)^2}{N^3} \sum_{n=1}^N \sigma^2 + \frac{1}{N^3} \sum_{n=1}^N \sum_{j \neq n} \sigma^2 \\
&= \frac{(N-1)^2}{N^2} \sigma^2 + \frac{N-1}{N^2} \sigma^2 \\
&= \frac{N^2 - 2N + 1 - 1 + N}{N^2} \sigma^2 \\
&= \frac{N-1}{N} \sigma^2.
\end{aligned}$$

We see that the ML estimate of the variance is biased (but asymptotically as $N \rightarrow \infty$ it is unbiased).