10-606: **Mathematical Foundations for Machine Learning** **Fall 2021**

Recitation 6: MLE, Expected Value, and Variance Exercises

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

6.1 Gaussian MLE

Recall that the pdf for a Gaussian random variable X is

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

Assume we have data $\mathcal{D} = \{x^{(i)}\}_{i=1}^N$ and we assume that $x^{(i)} \in \mathbb{R}$ and are i.i.d from $\mathcal{N}(\mu,\sigma)$. In the following we will work through deriving the MLE for both μ and σ.

Question 1a) Write the expression for the likelihood of the data with respect to parameters μ and σ.

Solution.

$$\begin{aligned}
\mathcal{L}(\mu,\sigma;\mathcal{D}) &= f(\mathcal{D};\mu,\sigma) = f(x^{(1)}, x^{(2)}, \dots, x^{(N)};\mu,\sigma) \\
&= \prod_{i=1}^N f(x^{(i)};\mu,\sigma) \\
&= \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x^{(i)}-\mu}{\sigma}\right)^2}
\end{aligned}$$

Question 1b) Write the expression for the log-likelihood of the data with parameters μ and σ.

Solution.

$$\begin{aligned}
\ell(\mu,\sigma;\mathcal{D}) &= \log \mathcal{L}(\mu,\sigma;\mathcal{D}) \\
&= \log \prod_{i=1}^N \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x^{(i)}-\mu}{\sigma}\right)^2} \\
&= \sum_{i=1}^N \log \left[\frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x^{(i)}-\mu}{\sigma}\right)^2}\right] \\
&= \sum_{i=1}^N \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \frac{-1}{2}\left(\frac{x^{(i)}-\mu}{\sigma}\right)^2 \\
&= -N\log(\sigma) - \frac{N}{2}\log(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^N (x^{(i)}-\mu)^2
\end{aligned}$$

Question 1c) Give the equation for the MLE of μ in terms of the data. Take the partial derivative with respect to μ, set to 0, and solve.

Solution.

$$\begin{aligned}\frac{\partial}{\partial\mu}\ell(\mu,\sigma;\mathcal{D}) &= \frac{-1}{\sigma^2}\sum_{i=1}^{N}(x^{(i)}-\mu)\\ &= \frac{1}{\sigma^2}\left(N\mu-\sum_{i=1}^{N}x^{(i)}\right)\end{aligned}$$

Setting to 0 and solving...

$$\begin{aligned}\frac{1}{\sigma^2}\left(N\mu-\sum_{i=1}^{N}x^{(i)}\right) &= 0\\ N\mu &= \sum_{i=1}^{N}x^{(i)}\\ \mu &= \frac{1}{N}\sum_{i=1}^{N}x^{(i)}\end{aligned}$$

Question 1d) Give the equation for the MLE of σ in terms of the data. Take the partial derivative with respect to σ, set to 0, and solve.

Solution.

$$\frac{\partial}{\partial\sigma}\ell(\mu,\sigma;\mathcal{D}) = \frac{-N}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{N}(x^{(i)}-\mu)^2$$

Setting to 0 and solving...

$$\begin{aligned}\frac{-N}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{N}(x^{(i)}-\mu)^2 &= 0\\ \frac{1}{\sigma^3}\sum_{i=1}^{N}(x^{(i)}-\mu)^2 &= \frac{N}{\sigma}\\ \sigma^2 &= \frac{1}{N}\sum_{i=1}^{N}(x^{(i)}-\mu)^2\\ \sigma &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x^{(i)}-\mu)^2}\end{aligned}$$

6.2 Practice with Expectation, Variance, and Covariance

6.2.1 Loaded Dice

You happen to have a loaded die where the faces of the die do not come up with equal probability. Let X be the random variable for the value of the die after being rolled. You know that,

$$P(X = 1) = 0.2$$
$$P(X = 2) = 0.05$$
$$P(X = 3) = 0$$
$$P(X = 4) = 0.25$$
$$P(X = 5) = 0.3$$
$$P(X = 6) = 0.2$$

It might be useful to use python for the following...

Question 2.1 a) Calculate the expected value for the die roll. $\mathbb{E}[X] = \ldots$

Solution.

$$\mathbb{E}[X] = 1 \cdot 0.2 + 2 \cdot 0.05 + 3 \cdot 0 + 4 \cdot 0.25 + 5 \cdot 0.3 + 6 \cdot 0.2 = 4$$

Question 2.1 b) Calculate the variance for the die roll. $\mathrm{Var}(X) = \ldots$

Solution.

$$\begin{aligned}\mathrm{Var}(X) =& (1-4)^2 \cdot 0.2 + (2-4)^2 \cdot 0.05 + (3-4)^2 \cdot 0 \\ &+ (4-4)^2 \cdot 0.25 + (5-4)^2 \cdot 0.3 + (6-4)^2 \cdot 0.2 \\ &= 3.1\end{aligned}$$

6.2.2 Mean and Variance of Bernoulli

Let $X \sim Bernoulli(q)$, i.e. $P(X = 1) = q$ and $P(X = 0) = 1 - q$.

Question 2.2 a) Find $\mathbb{E}[X]$.

Solution.

$$\mathbb{E}[X] = 1 \cdot p + 0 \cdot (1 - p) = p$$

Question 2.2 b) Show that $\text{Var}(X) = q(1-q)$.

Solution.

$$\begin{aligned}\text{Var}(X) &= (1-\mathbb{E}[X])^2 \cdot q + \mathbb{E}[X]^2 \cdot (1-q) \\ &= q(1-q)^2 + q^2(1-q) \\ &= q(1-q)(1-q+q) \\ &= q(1-q)\end{aligned}$$

6.2.3 Coin Flipping

Your friend claims that they have two coins that are magically linked. In particular, they claim that the outcome of the first coin, represented by random variable X_1, influences the result of the outcome of the second coin, represented by random variable X_2. In other words, the claim is that the two random variables are not independent.

x_1	x_1	$P(X_1 = x_1, X_2 = x_2)$
0	0	0.1
0	1	0.2
1	0	0
1	1	0.7

Question 2.3 a) Compute the expected values $\mathbb{E}[X_1]$ and $\mathbb{E}[X_2]$.

Solution.

$$\begin{aligned}\mathbb{E}[X_1] &= 0 \cdot 0.1 + 0 \cdot 0.2 + 1 \cdot 0 + 1 \cdot 0.7 = 0.7 \\ \mathbb{E}[X_2] &= 0 \cdot 0.1 + 1 \cdot 0.2 + 0 \cdot 0 + 1 \cdot 0.7 = 0.9\end{aligned}$$

Question 2.3 b) Recall that

$$\text{Cov}(X_1, X_2) = \mathbb{E}\left[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])\right]$$

Compute the covariance between the two coins, $\text{Cov}(X_1, X_2)$.

Solution.

$$\begin{aligned}\text{Covar}(X_1, X_2) =& (0.1)(0-0.7)(0-0.9) \\ &+ (0.2)(0-0.7)(1-0.9) \\ &+ (0)(1-0.7)(0-0.9) \\ &+ (0.7)(1-0.7)(1-0.9) \\ &= 0.07\end{aligned}$$

Question 2.3 c) Consider two non-magic coins $Y_1 \sim Bernoulli(q_1)$ and $Y_2 \sim Bernoulli(q_2)$. In particular, $Y_1 \perp\!\!\!\perp Y_2$. Show that $\text{Covar}(Y_1, Y_2) = 0$. **An important note!** If two random variables are independent their covariance is 0; but if the covariance of two random variables is 0, it does not necessarily mean they are independent.

Solution.

$$\begin{aligned}
\text{Covar}(Y_1, Y_2) &= \sum_{y_1 \in \{0,1\}} \sum_{y_2 \in \{0,1\}} (y_1 - \mathbb{E}[Y_1])(y_2 - \mathbb{E}[Y_2]) P(Y_1 = y_1, Y_2 = y_2) \\
&\overset{(independence)}{=} \sum_{y_1 \in \{0,1\}} \sum_{y_2 \in \{0,1\}} (y_1 - \mathbb{E}[Y_1])(y_2 - \mathbb{E}[Y_2]) P(Y_1 = y_1) P(Y_2 = y_2) \\
&= (0 - q_1)(0 - q_2)(1 - q_1)(1 - q_2) + (0 - q_1)(1 - q_2)(1 - q_1)q_2 \\
&\quad + (1 - q_1)(0 - q_2)q_1(1 - q_2) + (1 - q_1)(1 - q_2)q_1 q_2 \\
&= q_1 q_2(1 - q_1)(1 - q_2) - q_1 q_2(1 - q_2)(1 - q_1) - q_1 q_2(1 - q_1)(1 - q_2) + q_1 q_2(1 - q_1)(1 - q_2) \\
&= 0
\end{aligned}$$

6.3 Some More Probability Practice

Consider random variables A, B, C. For all of the following questions, you can assume that $A \perp\!\!\!\perp B$ but you cannot assume that $AB|C$.

Question 3.1 Write $P(A, B, C)$ in terms of $P(A), P(B), P(C|A, B)$.

$$P(A, B, C) = P(C|A, B)P(A, B) = P(C|A, B)P(A)P(B)$$

Question 3.2 Write $P(A, B|C)$ in terms of $P(A), P(B), P(C)$, and $P(C|A, B)$.

Solution.

$$\begin{aligned}
P(A, B|C) &= \frac{P(A, B, C)}{P(C)} \\
&= \frac{P(C|A, B)P(A, B)}{P(C)} \\
&= \frac{P(C|A, B)P(A)P(B))}{P(C)}
\end{aligned}$$

Question 3.3 Write $P(B, C|A)$ in terms of $P(B|C, A)$ and $P(C|A)$.

Solution.

$$P(B, C|A) \overset{(ChainRule)}{=} P(B|C, A)P(C|A)$$

Question 3.4 Write $P(C)$ in terms of $P(A), P(B)$, and $P(C|A, B)$. Note that your final answer might not have, for example, $P(A)$ explicitly in the answer; it will likely appear as $P(A = a)$.

Solution.

$$\begin{aligned}
P(C) &= \sum_a \sum_b P(C, A = a, B = b) \\
&= \sum_a \sum_b P(C|A = a, B = b)P(A = a, B = b) \\
&= \sum_a \sum_b P(C|A = a, B = b)P(A = a)P(B = b)
\end{aligned}$$

Question 3.5 Write $P(B|C)$ in terms of $P(A), P(B)$, and $P(C|A, B)$.

Solution.

$$\begin{aligned}
P(B|C) &= \frac{P(B, C)}{P(C)} \\
&= \frac{\sum_a P(B, C, A = a)}{\sum_a \sum_b P(C, A = a, B = b)} \\
&= \frac{\sum_a P(C|B, A = a)P(A = a, B)}{\sum_a \sum_b P(C|A = a, B = b)P(A = a, B = b)} \\
&= \frac{\sum_a P(C|B, A = a)P(A = a)P(B)}{\sum_a \sum_b P(C|A = a, B = b)P(A = a)P(B = b)}
\end{aligned}$$