Lisa Yan
CS109

Section #7 Warmup Solutions

Based on the work of many CS109 staffs

1 Lecture 20, 5-20-20: Parameters and MLE

Suppose $x_1, \ldots, x_n$ are iid samples from some distribution with density function $f_X(x; \theta)$, where θ is unknown. Recall that the likelihood of the data is

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

Recall we solve an optimization problem to find $\hat{\theta}$ which maximizes L.

1. Write an expression for the log-likelihood, $LL(\theta) = \log L(\theta)$.
2. Why can we optimize $LL(\theta)$ rather than $L(\theta)$?
3. Why do we optimize $LL(\theta)$ rather than $L(\theta)$?

1. $LL(\theta) = \sum_{i=1}^{n} \log f_X(x_i; \theta)$
2. Logarithms are monotonic. This means that if $f(a) > f(b)$, then $\log(f(a)) > \log(f(b))$, so correctness of arg max is preserved.
3. Logs turn products into sums, which makes taking the derivative much simpler.

2 Lecture 21, 5-22-20: Beta

1. Suppose you have a coin where you have no prior belief on its true probability of heads p. How can you model this belief as a beta distribution?
2. Suppose you have a coin which you believe is fair, with "strength" α. That is, pretend you've seen α heads and α tails. How can you model this belief as a Beta distribution?
3. Now suppose you take the coin from the previous part and flip it 10 times. You see 8 heads and 2 tails. How can you model your posterior belief of the coin's probability of heads?

1. $Beta(1, 1)$ is a uniform prior.
2. $Beta(\alpha + 1, \alpha + 1)$. This is our prior belief about the distribution.
3. $Beta(\alpha + 9, \alpha + 3)$