Homework - October 1, 2024

Student: Ayrton Chilibeck, achilibe@ualberta.ca
Lecturer: Lili Mou, UoA.F24.466566@gmail.com

> **Problem 1: Bernoulli Distribution**
>
> Consider tossing a coin for a few times, where the outcomes follow a Bernoulli distribution parametrized by π. In other words, the probability of having a head is π, whereas the probability of having a tail is $1 - \pi$.
> Suppose we have N_1 heads and N_2 tails. Give the formula for the likelihood of π. Show that the negative log-likelihood is convex in π.

Recall that the likelihood function $\mathcal{L}$ is the same as the probability density function. So for a Bernoulli distribution we have

$$L(\pi) = \pi^{N_1}(1-\pi)^{N_2}$$

Which is the joint probability of observing the exact number of heads and tails we obtained in our trial.

We can show the negative-log-likelihood function is convex through the definition of concavity or more simply through a second derivative test:

$$\begin{aligned} -\log L = \mathcal{L} &= -\left(N_1 \log(\pi) + N_2 \log(1-\pi)\right) \\ \frac{d\mathcal{L}}{d\pi} &= -\frac{N_1}{\pi} + \frac{N_2}{1-\pi} \\ \frac{d^2\mathcal{L}}{d\pi^2} &= \frac{N_1}{\pi^2} + \frac{N_2}{(1-\pi)^2} \end{aligned}$$

Both terms are positive in the range $\pi \in (0,1)$, so the function is indeed convex. We can also prove this using the definition of convexity.

> **Problem 2: Constraints**
>
> Consider the training objective $J = ||Xw - t||^2$ with a constraint $||w||^2 \leq C$ for some constant C. How would the hypothesis class capacity, overfitting/underfitting and bias/variance vary according to C?

	Larger C	Smaller C
Model Capacity	Larger	Smaller
Overfitting/Underfitting	Overfitting	Underfitting
Bias	Low	High
Variance	High	Low