AI4Good: Solutions

June 2020

1 Maximum likelihood estimation

1. (a) Answer: $y_i \sim \mathcal{N}(f_\theta(x_i), \sigma^2)$
   Explanation: By considering the cumulative distribution function of ϵ_i, we can show that y_i has the same distribution as ϵ_i with the mean shifted by $f_\theta(x_i)$ (a constant).

   (b) *likelihood*$(\theta) = P[y_1, \ldots, y_n | x_1, \ldots x_n, \theta]$
   Use the independence of $y_1, \ldots, y_n$ and the definition of the Gaussian distribution to derive the final form.
   In order to compute the logarithm, use the property:
   $\log(ab) = \log(a) + \log(b)$
   Answer: *loglikelihood*$(\theta) = \cdots = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - f_\theta(x_i))^2$

   (c) Answer: The second term corresponds to the Mean Squared Error loss function. Maximizing the likelihood is then equivalent to minimizing the loss.

2. (a) Answer: *loglikelihood*$(\theta) = -\frac{1}{2}\sum_{i=1}^{n}\log(2\pi\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{y_i - f_\theta(x_i))^2}{\sigma_i^2}$

   (b) The errors are not identically distributed. This means that not all errors contribute equally: a big error when the variance (σ_i^2) is small decreases the likelihood much more than a big error when variance is already large.

2 Bayesian analysis and biased coin

1. Likelihood: *likelihood*$(\theta) = P[HHH|\theta] = \theta^3$

   MLE: $\arg\max_\theta$ *likelihood*(θ)

   Since we know that $\theta \in [0, 1]$, we can plot the function θ^3 in order to find the maximum in the $[0, 1]$ range.

   Answer: Likelihood is maximized for $\theta^* = 1$, i.e. if the coin toss always results in H. It is a very naive estimator in this case since we know that such coin would be unlikely.

2. Mean and mode

   Beta distribution:

   $Beta(\alpha, \beta) = \theta^{\alpha+2}(1-\theta)^{\beta-1}$

   In order to find the mean and mode, we first compute posterior distribution. We know that the prior $p[\theta|\alpha, \beta]$ is parameterised by Beta distribution.

   From Bayes rule:

   $p(\theta|HHH; \alpha, \beta) = \frac{p[HHH|\theta;\alpha,\beta]p[\theta|\alpha,\beta]}{p[HHH|\alpha,\beta]} = \cdots \sim \theta^3\theta^{\alpha+2}(1-\theta)^{\beta-1} \sim Beta(\alpha+3, \beta)$

   Then we can check the definition of a mode and mean of a Beta distribution and insert the parameter values of $\alpha + 3, \beta$.

   Answer:

   $mode = \frac{\alpha+2}{\alpha+\beta+1}$ $mean = \frac{\alpha+3}{\alpha+\beta+3}$

   Maximum likelihood estimator does not include prior knowledge about the world. In this case, MLE returned $\theta^* = 1$ and maximizing posterior (which includes the prior belief that a probability of getting a head in a coin toss is roughly $\theta = \frac{1}{2}$) gives an answer of $\theta^* \sim \frac{1}{2}$.