10 Machine Learning and Bayesian Inference (sbh11)

The central limit theorem tells us that, if X_i are random variables, μ is the mean of X_i and σ^2 is the variance of X_i then, under suitable conditions

$$\frac{\hat{X}_n - \mu}{\sigma/\sqrt{n}} \to N(0,1)$$

where $N(0,1)$ denotes the normal density with mean 0 and variance 1, and

$$\hat{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

(*a*) Let Y have density $N(0,1)$. We know that, for a parameter p, there is a constant z_p such that

$$\Pr(-z_p \leq Y \leq z_p) > p.$$

Show that with probability at least p, the quantity μ as defined above is in the interval described by $\hat{X}_n \pm z_p(\sigma/\sqrt{n})$. [3 marks]

(*b*) The quantity $\hat{X}_n$ can be regarded as an estimate of μ. If the random variables X_i take values in $\{0,1\}$, and are also independent and identically distributed, explain why it might make sense to estimate σ^2 as

$$\sigma^2 \simeq s = \hat{X}_n(1 - \hat{X}_n).$$

[3 marks]

(*c*) Define what it means for an estimate such as that suggested in Part (*b*) to be *unbiased*. Is the estimate suggested in Part (*b*) unbiased? Provide a proof of your answer. [7 marks]

(*d*) We have two binary classifiers h_1 and h_2, and a test set $\mathbf{s}$ containing 1000 examples. During testing, h_1 makes 105 errors and h_2 makes 120 errors. Explain how we can estimate a confidence interval of the kind defined in Part (*a*) for the difference $(\text{er}(h_1) - \text{er}(h_2))$ between the true error probabilities $\text{er}(h_1)$ and $\text{er}(h_2)$ of the classifiers. Your answer should be careful to state any assumptions or approximations being made. [7 marks]