DIS9

1 Surprise and Entropy

Part (a)

0.

Part (b)

∞.

Part (c)

Minimized when all the balls are black or while. In this case the entropy is 0. The entropy is maximized when half the balls are black and half the balls are white, in which case the entropy is 1.

Part (d)

The function is strictly concave. Note that the function $-x \log x$ is strictly concave in $[0, 1]$, and a sum of strictly concave functions is strictly concave.

2 Ensemble Learning

Part (a)

$$\mathbb{E}[Y] = \mu, \mathrm{Var}[Y] = \frac{1}{n}\sigma^2$$

Part (b)

For a sample point, the probability of not being chosen is $(1 - \frac{1}{n})^n$, which goes to $1/e \approx 0.37$ as n goes to ∞.

Part (c)

$$Var[Z] = Var\left[\frac{1}{n}\sum_{i=1}^{n} Z_i\right] = \frac{1}{n^2}\left(\sum_{i=1}^{n} Var[Z_i] + \sum\sum_{i\neq j} Cov[Z_i, Z_j]\right)$$

which simplifies to $\frac{\sigma^2-\rho}{n} + \rho$. We can see that for correlated random variables, the benefit of averaging is limited.

3 Decision Trees and Random Forests

Part (c)

Entropy always non-negative.

Part (d)

$$(1 - \frac{1}{d})^{Th}$$

Part (e)

$$(1 - \frac{1}{n})^{Tn}$$

Part (f)

$(1/2)^{40}$ and $(1/2)^{20}$.

It is very unlikely that a certain feature or sample point is never considered.