Name: Student ID:

ENSEMBLE LEARNING PROBLEM SHEET

1 These questions have both appeared in past examinations.

(a) If $\{X_i | i = 1, 2, \ldots, n\}$ is a set of correlated random variables such that

$$\mathbb{E}\left[X_i\right] = \mu \qquad \mathbb{E}\left[(X_i - \mu)(X_j - \mu)\right] = \begin{cases} \sigma^2 & \text{if } i = j \\ \rho\sigma^2 & \text{if } i \neq j \end{cases}$$

show

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] = \rho\sigma^2 + \frac{(1-\rho)\sigma^2}{n}$$

[10 marks]

$\overline{}$
10

(b) We consider a regression problem where the data $(\boldsymbol{x}, y)$ is distributed according to $\gamma(\boldsymbol{x}, y)$. We consider a learning machine that makes a prediction $\hat{f}(\boldsymbol{x}|\boldsymbol{\theta})$, where the parameters, $\boldsymbol{\theta}$ are trained using a stochastic algorithm that returns parameters distributed according to a probability density $\rho(\boldsymbol{\theta})$. We can define the mean machine as $\hat{m}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\theta}\sim\rho}\left[\hat{f}(\boldsymbol{x}|\boldsymbol{\theta})\right]$. We assume that

$$\mathbb{E}_{(\boldsymbol{x},y)\sim\gamma}\left[(\hat{m}(\boldsymbol{x}) - y)^2\right] = B, \qquad \mathbb{E}_{(\boldsymbol{x},y)\sim\gamma}\left[\mathbb{E}_{\boldsymbol{\theta}\sim\rho}\left[\left(\hat{f}(\boldsymbol{x}|\boldsymbol{\theta}) - \hat{m}(\boldsymbol{x})\right)^2\right]\right] = V.$$

That is, we can define a bias B and variance V. We now consider ensembling n machines

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \hat{f}(\boldsymbol{x}|\boldsymbol{\theta}_i)$$

where $\boldsymbol{\theta}_i$ are drawn independently from $\rho(\boldsymbol{\theta})$. Compute the expected generalisation error of $\hat{f}_n(\boldsymbol{x})$. (Note this is different from the usual bias-variance calculation because we are averaging the performance of n machines). [10 marks]

$\overline{}$
10

End of question 1

(a) $\frac{}{10}$ (b) $\frac{}{10}$ Total $\frac{}{20}$

