

### Exercise 1 - Variance

Show that for two independent random variables,  $X, Y$  and arbitrary  $a, b \in \mathbb{R}$ , the following equality holds

$$\text{Var}(aX + bY) = a^2 \cdot \text{Var}(X) + b^2 \cdot \text{Var}(Y).$$

### Exercise 2 - Variance / Bias Decomposition

Let  $D = \{(x_i, y_i) | i = 1 \dots n\}$  be a dataset obtained from the true underlying data distribution  $P$ , i.e.  $D \sim P^n$ . And let  $h_D(\cdot)$  be a classifier trained on  $D$ . Show the variance bias decomposition

$$\underbrace{\mathbb{E}_{D,x,y} [(h_D(x) - y)^2]}_{\text{Expected test error}} = \underbrace{\mathbb{E}_{D,x} [(h_D(x) - \hat{h}(x))^2]}_{\text{Variance}} + \underbrace{\mathbb{E}_{x,y} [(\hat{y}(x) - y)^2]}_{\text{Noise}} + \underbrace{\mathbb{E}_x [(\hat{h}(x) - \hat{y}(x))^2]}_{\text{Bias}^2}$$

where  $\hat{h}(x) = \mathbb{E}_{D \sim P^n} [h_D(x)]$  is the expected regressor over possible training sets, given the learning algorithm  $\mathcal{A}$  and  $\hat{y}(x) = \mathbb{E}_{y|x} [y]$  is the expected label given  $x$ . As mentioned in the lecture, labels might not be deterministic given  $x$ . To carry out the proof, proceed in the following steps:

(a) Show that the following identity holds

$$\mathbb{E}_{D,x,y} [(h_D(x) - y)^2] = \mathbb{E}_{D,x} [(\hat{h}_D(x) - \hat{h}(x))^2] + \mathbb{E}_{x,y} [(\hat{h}(x) - y)^2]. \quad (1)$$

(b) Next, show

$$E_{x,y} [(\hat{h}(x) - y)^2] = E_{x,y} [(\hat{y}(x) - y)^2] + E_x [(\hat{h}(x) - \hat{y}(x))^2] \quad (2)$$

which completes the proof by substituting (2) into (1).

### Exercise 3 - Ensembling

Download the file `exercises06-ensembling.ipynb` from quercus. It contains basic Pytorch code training a classifier on MNIST. Modify that code such that it trains an ensemble of 5-10 neural networks and computes their average prediction once trained.