Exercise 1 - Variance

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

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

Exercise 2 - Variance / Bias Decomposistion

Let $D = \{(x_i, y_i)|i = 1 \ldots n\}$ 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

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

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

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

(b) Next, show

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

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.