Exercise 12-1  Bias vs. Variance
Consider the following learning curves showing training and cross-validation error for two different models when the respective model is trained on an increasing amount of data. For both cases, discuss bias and variance. What are indicators for a high variance or a high bias problem?

Exercise 12-2  Bias-Variance Decomposition
Assume a fixed distribution $P(x)$ over $x$ and a dependent variable $y = f(x) + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. We try to model this relationship by a function $\hat{f}(x, w)$ whose parameter $w$ we learn. For short, we write $f(x) = f$ and $\hat{f}(x, w) = \hat{f}$. Consider the expected square loss $\mathbb{E}[L] = \mathbb{E}[(\hat{f} - y)^2]$.

(a) Show that: $\mathbb{E}[(\hat{f} - y)^2] = \mathbb{E}[(\hat{f} - f)^2] + \mathbb{E}[(f - y)^2]$
   with $\mathbb{E}[(f - y)^2] = \text{Var}[y] = \sigma^2$ being the intrinsic noise of the data.
   Hint: Make use of the calculation rules for the expected value: For two random variables $X$ and $Y$ and a constant $c$ it holds that: $\mathbb{E}[cX + Y] = c\mathbb{E}[X] + \mathbb{E}[Y]$. Moreover, if $X$ and $Y$ are independent: $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.

(b) Now show that: $\mathbb{E}[(\hat{f} - f)^2] = \mathbb{E}[(\hat{f} - \mathbb{E}[\hat{f}])^2] + (\mathbb{E}[\hat{f}] - f)^2$
   where $\mathbb{E}[(\hat{f} - \mathbb{E}[\hat{f}])^2]$ is the variance and $(\mathbb{E}[\hat{f}] - f)^2$ the bias of our estimation $\hat{f}$.
   Hint: Subtract and add $\mathbb{E}[\hat{f}]$ to the squared term in the loss and simplify.

(c) With that we have:
\[
\mathbb{E}[L] = \mathbb{E}[(\hat{f} - y)^2] = \mathbb{E}[(\hat{f} - \mathbb{E}[\hat{f}])^2] + (\mathbb{E}[\hat{f}] - f)^2 + \mathbb{E}[(f - y)^2] = \text{Var}[\hat{f}] + (\text{Bias}(\hat{f}))^2 + \sigma^2.
\]
What is the minimum of the expected loss?