Learning Theory

- Using learning theory, we can make formal statements or give guarantees on:
  - Expected performance of a learning algorithm.
  - Number of examples required to attatin a certain level of test accuracy.
  - Hardness of learning problems in general.

Bias, Variance and Model Complexity

Def. Bias

The tendency to consistently learn the same wrong thing.

- The bias is error from erroneous assumptions in algorithm.
- High bias causes an algorithm to miss the relevant relations between features and outputs.

Def. Variance

The tendency to learn random things irrespective of the real signal.

- The variance is error from sensitivity to small fluctutations in the training set.
- High variance causes an algorithm to model the random noise rather than the outputs.

- Loss function for measuring errors between Y and $\hat{f}(X)$

$$L(Y, \hat{f}(X)) = \begin{cases} (Y - \hat{f}(X))^2, & \text{squared error} \\ |Y - \hat{f}(X)|, & \text{absolute error} \end{cases}$$

- Test / generalized error $\text{Err}_{\mathcal{D}} = \mathbb{E}[L(Y, \hat{f}(X)) \mid D]$, where $\mathcal{D}$ denotes the training set.
- Expected prediction / test error $\text{Err} = \mathbb{E}[L(Y, \hat{f}(X))] = \mathbb{E}[\text{Err}_{\mathcal{D}}]$.
- Training error $\overline{\text{err}} = \frac{1}{m} \sum_{i=1}^{m} L(y_i, \hat{f}(x_i))$

- Simple model have high bias and small variance.
- Complex models have small bias and high variance.
- The bad performance (low accuracy on test data) could be due to either high bias (underfitting) or high variance (overfitting).

- High bias: Both training and test error are large.
- High variance: Small training error, large test error.