Supervised Learning:

* Have both predictors and outcomes for each observation in training data
* Used to predict response variables
* Eg: How many units will be sold given a budget, how much sales will increase if advertising increased, which type of advertising is most successful

Unsupervised learning

* Have predictors, but no responses
* Eg: Using health device measurements, cluster observations into different user groups. Using customer complaints, identify main issues that are reported

Why estimate *f*

* Prediction
  + Ŷ = ̂f(x), where ̂f is a black box
* Inference
  + How Y is changing as a function of X

Supervised learning for prediction

* Use data (X1, Y1),…,(Xn,Yn) to estimate ̂f
* Use ̂f to estimate ̂Y for a new X:
  + ̂Y = ̂f (X)
* How accurately we can predict ̂Y?
  + Bias: Y - ̂Y = f(X) + - ̂f(X)
  + Reducible error: f(x) -
  + Irreducible error:

Bias-Variance Trade-Off

* The [bias](https://en.wikipedia.org/wiki/Bias_of_an_estimator) is an error from erroneous assumptions in the learning [algorithm](https://en.wikipedia.org/wiki/Algorithm). High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).
* The [variance](https://en.wikipedia.org/wiki/Variance) is an error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random [noise](https://en.wikipedia.org/wiki/Noise_(signal_processing)) in the training data, rather than the intended outputs ([overfitting](https://en.wikipedia.org/wiki/Overfitting)).
* As flexibility of increases, bias decreases, but variance increases
* Find a method where variance and bias is low