How does Supervised Learning Work?

Data: $\{x\_i, y\_i\}$
i=1, 2, …, n

Goal: learn w
y <—- $f(x; w)$

Minimize an objective function

$$F(w) = \sum_{i=1}^{n} \text{loss}(y_i, f(x_i; w))$$

$$\text{loss} = \left[y_i - f(x_i; w) \right]^2$$

$$\text{loss} = \mathbf{1}\left\{ y_i \neq f(x_i; w) \right\}$$

$$\text{loss} = \begin{cases} -\log f(x_i; w) & \text{if } y_i = 1 \\ -\log(1 - f(x_i; w)) & \text{if } y_i = 0 \end{cases}$$

How does Supervised Learning Work?

Data: $\{x\_i, y\_i\}$
i=1, 2, …, n

Goal: learn w
y <—- $f(x; w)$

Minimize an objective function

$$F(w) = \sum_{i=1}^{n} \text{loss}(y_i, f(x_i; w))$$

1. The minimizer w^* may be in closed form.
2. Try optimization algorithms that can guarantee to converge to the global minimizer.
3. In the worst case, try *gradient descent*.

Overview

- Types of statistical learning problems
- Why learning is difficult?
- Bias variance tradeoff
- An example: kNN *vs* Linear Regression (in a separate pdf file)
- Not all about prediction

Challenges

- Training error underestimates test/generalization error.
- Overfitting: perform well on the training data but not on the future (unseen) data.
- p denotes the number of parameters the regression/classification function f has, i.e., the number of parameters we need to learn from the data.
- The gap between the two errors (training *vs.* test) gets large when p is large.