Supervised Learning

Description

Given: a training set of data with (input, output) pairs generated by an unknown function f

Goal: create a hypothesis function h that gives good outputs for h(input)

- degrees of freedom: number of parameters of h
- discrete vs. continuous output: classification vs. regression
- pick good function: avoid underfitting or overfitting

Classification Idea

Given a labeled data set and unlabeled input, we want to produce a label

Rather than coming up with an estimated function, look at 'nearby' labeled data to guess label

How many neighbors should we look at?

k-Nearest Neighbors

Look at k nearest labeled neighbors, pick the majority
When $k = 1$, labels are Voronoi cells
Measuring distance: Euclidean or Manhattan

$$d(a,b) = \sqrt{\sum_i (a_i - b_i)^2}$$

normalize dimensions:

$$\text{for each dimension: } x_i' = \frac{x_i - xmin_i}{xmax - xmin}$$

- good: robust to noise, choosing k is easy
- bad: memory use, k-d tree, irrelevant features, high dimensions

Regression

Instead of producing a label, produce a numerical output with hypothesis function $\hat{y}$

$$\hat{y} = w_0 f_0(x) + w_1 f_1(x) + w_2 f_2(x) + \ldots$$

Examples:

$$\hat{y} = w_0 + w_1 x$$

$$\hat{y} = w_0 + w_1 x + w_2 x^2$$

$$\hat{y} = w_0 + w_1 \sin x$$

Regression

Instead of producing a label, produce a numerical output with hypothesis function $\hat{y}$

$$\hat{y} = w_0 f_0(x) + w_1 f_1(x) + w_2 f_2(x) + \ldots$$

Examples:

$$\hat{y} = w_0 + w_1 x$$
$$\hat{y} = w_0 + w_1 x + w_2 x^2$$
$$\hat{y} = w_0 + w_1 \sin x$$

Least Mean Squares Procedure

$$\hat{y} = wx$$

$$w \leftarrow w - \alpha(\hat{y} - y)x$$

For $x = \langle 1, x_1, x_2 \rangle$, updates are:

$$w_0 \leftarrow w_0 - \alpha(\hat{y} - y)$$

$$w_1 \leftarrow w_1 - \alpha(\hat{y} - y)x_1$$

$$w_2 \leftarrow w_2 - \alpha(\hat{y} - y)x_2$$

$\alpha \approx 1/N$? Or $100/(100+N)$? or?

Least Mean Squares Procedure

$$\hat{y} = wx$$
$$w \leftarrow w - \alpha(\hat{y} - y)x$$

For $x = \langle 1, x_1, x_2 \rangle$, updates are:

$$w_0 \leftarrow w_0 - \alpha(\hat{y} - y)$$
$$w_1 \leftarrow w_1 - \alpha(\hat{y} - y)x_1$$
$$w_2 \leftarrow w_2 - \alpha(\hat{y} - y)x_2$$

$\alpha \approx 1/N$? Or $100/(100+N)$? or?