Robust Regression

February 5, 2015

Let $X = \mathbb{R}^D$ and $Y = \mathbb{R}$. Let $T = \{(x_1, y_1), \ldots, (x_N, y_N)\}$ be a set of N training examples, with $x_i \in X$ and $y_i \in Y$. We would like to use T to estimate a function $f : X \to Y$.

In a generalized linear model we consider functions f that are linear combinations of a fixed set of basis functions. We have M basis functions $\phi_i : X \to \mathbb{R}$ for $1 \leq i \leq M$. Let $\phi(x) = (\phi_i(x), \ldots, \phi_M(x)) \in \mathbb{R}^M$. A vector of coeffcients $w \in \mathbb{R}^M$ defines a function $f_w(x) = w^T \phi(x)$.

In least squares (LSQ) regression we select w minimizing a sum of squared differences

$$E(w) = \sum_{i=i}^{N} (f_w(x_i) - y_i)^2$$

This can be done by solving a linear system. But the approach is not robust to outliers.

In least absolute deviation (LAD) regression we select w minimizing a sum of absolute values

$$Q(w) = \sum_{i=i}^{N} |f_w(x_i) - y_i|$$

As we discussed in class this objective is robust to outliers.

(The relationship between LSQ and LAD is similar to the relationship between the mean and median of real values.)

The problem of fitting a function by minimizing $Q(w)$ can be reduced to a linear programming problem as follows.

Let $z = (w_1, \ldots, w_M, e_1, \ldots, e_N) \in \mathbb{R}^{M+N}$ be a set of variables representing the cofficients we want to estimate and the errors of f_w on each training example.

The objective $Q(w)$ can be expressed as a linear function of z

$$Q(w) = c^T z$$

where

$$c = (\underbrace{0, \ldots, 0}_{\text{M entries}}, \underbrace{1, \ldots, 1}_{\text{N entries}})$$

To enforce that $e_i = |f_w(x_i) - y_i|$ we use the constraints

$$e_i \geq f_w(x_i) - y_i$$

$$e_i \geq y_i - f_w(x_i)$$

Together these two constraints are equivalent to requiring that

$$e_i \geq |f_w(x_i) - y_i|$$

Since we are minimizing $c^T z = \sum_{i=1}^{N} e_i$ we would never select e_i to be greater than it needs to be. Therefore we will have $e_i = |f_w(x_i) - y_i|$.

In summary LAD regression involves solving a linear program with $N + M$ variables and $2N$ linear constraints:

$$\min_z c^T z$$

subject to 2 constraints for each $1 \leq i \leq N$,

$$e_i \geq w^T \phi(x_i) - y_i$$

$$e_i \geq y_i - w^T \phi(x_i)$$

Both of these constraints are linear in z. We can re-write them in the form $a^T z \leq b$,

$$(\phi(x_i), 0, \ldots, 0, -1, 0, \ldots, 0)^T z \leq y_i$$

$$(-\phi(x_i), 0, \ldots, 0, -1, 0, \ldots, 0)^T z \leq -y_i$$

where the -1 value is in the position that will multiply e_i within z.