Multi-Dimensional Regression

- More elegant/general to do $\nabla_{\mathbf{w}} R = 0$ with linear algebra
- Rewrite empirical risk in vector-matrix notation:
- Can add more dimensions by adding columns to X matrix and rows to w vector:

$$
\begin{aligned}
R(\mathbf{w}) &= \frac{1}{2N}\sum_{i=1}^{N}(y_i - w_1 x_i - w_0)^2 \\
&= \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - [1, x_i]\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}\right)^2 \\
&= \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - [1, x_i(1) \ldots x_i(d)]\begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}\right)^2 \\
&= \frac{1}{2N}\left\| \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} 1 & \mathbf{x}_1 \\ \vdots & \vdots \\ 1 & \mathbf{x}_N \end{bmatrix}\begin{bmatrix} w_0 \\ \mathbf{w} \end{bmatrix} \right\|^2 \\
&= \frac{1}{2N}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2
\end{aligned}
$$

Multi-Dimensional Regression

- Solving gradient=0

$$\nabla_\theta R = 0$$

$$\nabla_\theta \left(\frac{1}{2N} \left\| \mathbf{y} - \mathbf{X}\theta \right\|^2 \right) = 0$$

$$\frac{1}{2N} \nabla_\theta \left(\left(\mathbf{y} - \mathbf{X}\theta \right)^T \left(\mathbf{y} - \mathbf{X}\theta \right) \right) = 0$$

$$\frac{1}{2N} \nabla_\theta \left(\mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X}\theta + \theta^T \mathbf{X}^T \mathbf{X}\theta \right) = 0$$

$$\frac{1}{2N} \left(-2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\theta \right) = 0$$

$$\mathbf{X}^T \mathbf{X}\theta = \mathbf{X}^T \mathbf{y}$$

$$\theta^* = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y}$$

- In Matlab: "t=pinv(X)*y" or "t=X\y" or "t=inv(X'*X)*X'*y"

Multi-Dimensional Regression

- Solving gradient=0

$$\mathbf{X}^T\mathbf{X}\theta = \mathbf{X}^T\mathbf{y}$$

$$\theta^* = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

- In Matlab: "t=pinv(X)*y" or "t=X\y" or "t=inv(X'*X)*X'*y"
- If the matrix X is skinny, the solution is probably unique
- If X is fat (more dimensions than points) we get multiple solutions for theta which give zero error.

- The pseudeoinverse (pinv(X)) returns the theta with zero error and which has the smallest norm.

$$\min_\theta \left\|\theta\right\|^2 \ \textit{such that}\ \mathbf{X}\theta = \mathbf{y}$$

Radial Basis Functions

- Each training point leads to a bump function

$$f(\mathbf{x};\theta) = \sum_{k=1}^{N} \theta_k \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x} - \mathbf{x}_k\right\|^2\right) + \theta_0$$

- Reuse solution from linear regression: $\theta^* = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$
- Can view the data instead as Q, a big matrix of size N x N

$$\mathbf{Q} = \begin{bmatrix} \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_1 - \mathbf{x}_1\right\|^2\right) & \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_1 - \mathbf{x}_2\right\|^2\right) & \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_1 - \mathbf{x}_3\right\|^2\right) \\ \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_2 - \mathbf{x}_1\right\|^2\right) & \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_2 - \mathbf{x}_2\right\|^2\right) & \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_2 - \mathbf{x}_3\right\|^2\right) \\ \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_3 - \mathbf{x}_1\right\|^2\right) & \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_3 - \mathbf{x}_2\right\|^2\right) & \exp\left(-\tfrac{1}{2\sigma^2}\left\|\mathbf{x}_3 - \mathbf{x}_3\right\|^2\right) \end{bmatrix}$$

- In this setting, X is invertible ,solution is just $\theta^* = \mathbf{Q}^{-1}\mathbf{y}$