Lecture 2 Jan 21 th.

What can we do with regression?

- prediction
- fitting

What is the general process of building/training a ml algorithm?

- collect data (care about bias and sample size)

Non-linear Regression//

Basis Function Regression//

Regression Function: $y = f(x) = \sum_k w_k b_k(x)$

where w_k is a weight or the parameters

$b_k(x)$ is a "simple" func. or the basis funcs.

y is the approximation of the relationship

Ex. The regre. function of linear regression is $y = f(x) = \sum_k w_k x$

The regre. function of polynomial regression of order 4 is:

$$y = f(x) = w_4 x^4 + w_3 x^3 + w_2 x^2 + w_1 x + w_0$$

Here the weights are $w_4 \cdots w_0$, the basis funcs. are $x^4 \cdots x^0$

We can rewrite above as

$$y = f(x) = \vec{b}(x)^T \vec{w}$$

$$= [\, b_1(x) \; b_2(x) \cdots b_n(x) \,] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

Radial Basis Functions (RBFs)//

A B.F.R. with basis function

$$b_i(x) = \exp\left(-\frac{(x-c_i)^2}{2\sigma^2}\right)$$

Thus $$f(x) = \sum_k w_k b_k(x)$$

$$= \sum_k w_k \exp\left(\frac{-(x-c_k)^2}{2\sigma^2}\right)$$

Where c_k is the center and σ^2 determines the width of the basis function. (These must be estimated)

To pick the centers, there are 3 methods

1. Place centers uniformly throughout the data. (can cause impractical number of centers in high dimensions)
2. Place centers for each data point. (This used more often since there's a limit for # of centers. (can be expensive for large sample size)
3. Cluster the data and place a center for each cluster.

To pick the width,

1. Trial and error
2. Use the average square distances to neighbouring centers, scaled by a constant. This allows you to use different widths for each basis function.

Calculation of Error of B.F.R//

We find the error using the least squares method.

$$E(w) = \sum_i (y_i - f(x_i))^2 = \sum_i (y - \sum_k w_k b_k(x))^2$$

$$= \| \vec{y} - B\vec{w} \|^2 \quad , \quad \vec{y} \in \mathbb{R}^n, \vec{w} \in \mathbb{R}^m, B \in \mathbb{R}^{n \times m}$$

Which is the same form as $\| \vec{y} - X\vec{w} \|^2$ which is the same form as linear regression. Thus,

$$w^* = B^{\dagger}\vec{y}$$

Issues with Overfitting //

Directly minimizing the square error can lead to overfitting; when the model fits too well with data and can not generalize.

When does this occur:

1. Problem not sufficiently constrained
2. More parameters than data
3. Fitting noise
4. Discarding uncertainty.

There are 2 ways to fix overfitting:

1. Add prior knowledge
2. Handling uncertainty

Adding Prior Knowledge : Regularization

Ex: higher degree polynomial always wins out on cross-validation since it can always fit the data the best.

To avoid this, we can regularize the objective function.

$$E(w) = \underbrace{\|\vec{y} - B\vec{w}\|^2}_{\text{data term}} + \underbrace{\lambda \|\vec{w}\|^2}_{\text{smoothness term.}} \qquad \|\vec{w}\| \text{ called weight decay}$$

$$= (y - Bw)^T (y - Bw) + \lambda w^T w$$

$$= w^T (B^T B + \lambda I) \vec{w} - 2\vec{w}^T B^T \vec{y} + \vec{y}^T \vec{y}$$

Thus $w^* = (B^T B + \lambda I)^{-1} B^T \vec{y}$

Now the model will become more smooth but will not have the smallest variance.

Artificial Neural Networks.

The basis function chosen is

$$g(a) = \frac{1}{1 + e^{-a}} \qquad \text{(Sigmoid Function)}$$

The trick here is that

$$\vec{y} = f(\vec{x}) = \sum_j \vec{w}_j^{(1)} g\left(\sum_k w_{k,j}^{(2)} x_k + b_j \right) + \vec{b}$$

A linear regression with weights $W(w_{k,j})$ are applied to $\vec{x}$ and is put through a Sigmoid Function and the output is used as features of another linear regression.

(1), (2) are different values.

Ex. The 1-D Case

$$y = \sum_j w_j \, g(w_j x + b_j) + b$$

Knn Regression

This method does not need model, does not need fitting

We find the k closests neibouring points and their (possibly weighted) average.

Ex. Pixel color correction, temperature prediction.

$$y = \frac{1}{k} \sum_{i \in N_k(x)} y_i$$

Weighted Example

$$y = \frac{\sum_{i \in N_k(x)} w(x_i) y_i}{\sum_{i \in N_k(x)} w(x_i)}, \quad w(x_i) = e^{-\|x_i - x\|^2 / 2\sigma^2}$$

R.B.F.