Kernels

Learning objectives
Kernel definition and examples
RBF algorithm
Compare kernel regression with KNN and RBF

Lyle Ungar

What is a kernel?

- k(**x**,**y**)
  - Measures the *similarity* between a pair of points ***x*** and ***y***
  - Symmetric and positive definite
- **Example: Gaussian kernel**
  - $k(\mathbf{x},\mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/\sigma^2)$
- **Uses**
  - K-NN
  - RBF
  - Kernel regression

Kernel definition

A symmetric function $k(\mathbf{x}_i, \mathbf{x}_j)$: $\boldsymbol{X} \times \boldsymbol{X} \rightarrow \mathbb{R}$

is a positive definite kernel on $\boldsymbol{X}$ if

$$\Sigma_{i,j} c_i c_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

for all c_i c_j $\mathbf{x}_i$, $\mathbf{x}_j$

summed over any set of i,j pairs

What is a kernel?

- $k(\mathbf{x},\mathbf{y})$
  - Measures the *similarity* between a pair of points ***x*** and ***y***
  - Symmetric and positive definite
  - Often tested using a *Kernel Matrix,*
    - a PSD matrix **K** with elements $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$ from all pairs of rows of a matrix *X* of predictors
    - A *PSD matrix* has only non-negative singular values
- **Uses**
  - Anywhere you want to replace inner products $x_1^T x_2$ with inner products of $\phi(x_1)^T \phi(x_2) = k(x_1, x_2)$

How are kernels selected?

- **Linear kernel**
  - $k(x,y) = x^T y$
- **Gaussian kernel**
  - $k(x,y) = \exp(-||x - y||^2/\sigma^2)$
- **Quadratic kernel**
  - $k(x,y) = (x^T y)^2$ or $(x^T y + 1)^2$
- **Combinations and transformations of kernels**

Radial Basis Functions (RBFs)

1) Pick k basis function centers μ_j

2) *Let* $h(\mathbf{x}) = w_1\, \phi_1(\mathbf{x}) + w_2\, \phi_2(\mathbf{x}) + \ldots w_k\, \phi_k(\mathbf{x})$

where

$\phi_j(\mathbf{x}) = k(\mathbf{x}, \mu_j) = \exp(-||\mathbf{x} - \mu_j||_2^2 / C)$

3) Estimate w using linear regression

RBFs can do …

- **Use k < p basis vectors**
  - Dimensionality reduction
  - Good for high dimensional feature spaces
- **Use k > p basis vectors**
  - Increases the dimensionality
  - Can make a formerly nonlinear problem linear
- **Use k=n basis vectors**
  - We will use this to switch to a *dual* representation

How to find the kernel centers?

- **Pick random points**
  - Generally a bad idea
- **RBF: do k-means clustering and use the centers of the clusters**
  - Works great!
- **Use all n of the training data points as kernel centers**
  - Requires regularization
- **Estimate them: nonlinear regression**
  - A good initialization helps

Kernel Regression

$$\hat{y}(\mathbf{x}) = \frac{\sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)y_i}{\sum_{i=1}^{n} K(\mathbf{x},\mathbf{x}_i)}$$

https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Lectures.KernelRegression

Kernel classification

$$\hat{y}(\mathbf{x}) = sign(\sum_{i=1}^{n} K(\mathbf{x}, \mathbf{x}_i)y_i)$$

KNN vs Kernel regression

- When is k-NN better than kernel regression?
- When is kernel regression better than k-NN

Positive Semi-Definite (PSD)?

- $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is positive semi-definite?
- A'A is guaranteed positive semi-definite?
- A positive semi-definite matrix can have negative entries in it?
- The covariance matrix is PSD?

True or False?