Kernels and Kernel Regression

Lyle Ungar

Learning objectives
Kernel definition and examples
RBF algorithm (again)
Kernel regression
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**

- $k(x,y) = \exp\left(-\frac{||x - y||^2}{\sigma^2}\right) = \exp\left(-\frac{d(x, y)^2}{\sigma^2}\right)$

**Uses of kernels**

- “soft” K-NN
- RBF
- Kernel regression, SVMs
Kernel definition

A symmetric function $k(x_i, x_j): \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$ is a positive definite kernel on $\mathbf{X}$ if

$$\sum_{i,j} c_i c_j k(x_i, x_j) \geq 0$$
for all $c_i, c_j x_i, x_j$
summed over any set of i,j pairs

We won’t actually use this
What is a kernel?

• $k(x, y)$
  • Measures the similarity between a pair of points $x$ and $y$
  • Symmetric and positive semi-definite (PSD)
  • Often tested using a Kernel Matrix,
    • a PSD matrix $K$ with elements $K_{ij} = k(x_i, x_j)$ from all pairs of rows of a matrix $X$
    • A PSD matrix has only non-negative eigenvalues
Kernel examples

- **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 $\mu_j$ using k-means clustering

2) Let $h(x) = w_1 \phi_1(x) + w_2 \phi_2(x) + \ldots w_k \phi_k(x)$

where

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

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**
  - Switches to a *dual* representation
Kernel Regression

\[ \hat{y}(x) = \frac{\sum_{i=1}^{n} K(x, x_i)y_i}{\sum_{i=1}^{n} K(x, x_i)} \]


Kernel classification

\[ \hat{y}(x) = \text{sign}(\sum_{i=1}^{n} K(x, x_i)y_i) \quad y_i = -1, 1 \]
KNN vs Kernel regression

- When is k-NN better than kernel regression?
- When is kernel regression better than k-NN
A kernel $k(x,y)$

- Measures the *similarity* between a pair of points $x$ and $y$
- Symmetric and positive semi-definite
- Often tested using a *Kernel Matrix*,
  - a PSD matrix $K$ with elements $K_{ij} = k(x_i,x_j)$ from all pairs of rows of a matrix $X$ of predictors
  - A PSD matrix has only non-negative eigenvalues
Kernel matrix example

- Pick a matrix $X$
  
  $$
  \begin{pmatrix}
  1 & 2 \\
  3 & 4 \\
  5 & 6 \\
  \end{pmatrix}
  $$

- Compute $K_{ij} = k(x_i, x_j)$

- Test the eigenvalues

- What is $K$ for $X$ using the linear kernel?
How was my speed

A Slow
B Good
C Fast