

## Forces with KRR

Malthe Kjær Bisbo

November 19, 2017

### 1 Kernel Ridge Regression

Kernel Ridge Regression (KRR) is basically an upgraded version of regularized linear regression, which is made linear by the kernel trick  $x^T x' \rightarrow k(x, x')$ , where  $k(.,.)$  is the kernel.

#### Regularized linear regression

In regularized linear (least square) regression we seek the coefficients  $w$  that minimizes the cost function

$$\min_w \left( \sum_i^N (\vec{w}^T x_i - y_i)^2 + \lambda \sum_i^N w_i^2 \right)$$

or

$$\min_w ((X\vec{w} - \vec{y})^2 + \lambda \|\vec{w}\|^2)$$

Where  $X$  is the matrix with the  $x_i^T$ 's as it's rows. The solution is

$$\vec{w} = X^T (X X^T + \lambda I)^{-1} \vec{y} = X^T \vec{\alpha}$$

Where we define  $\vec{\alpha} = (X X^T + \lambda I)^{-1} \vec{y}$ . Predictions are then performed using

$$y' = \vec{w}^T x' = (X^T \vec{\alpha})^T x' = \vec{\alpha}^T X x'$$

**Result:** To train and use the regularized linear regression model as an ML-predictor We thus need two steps.

#### 1. Training:

Calculate  $\vec{\alpha}$  from training data.

$$\vec{\alpha} = (X X^T + \lambda I)^{-1} \vec{y}$$

#### 2. Prediction:

Predict new value  $y'$  using

$$y' = \vec{\alpha}^T X x'$$

### Kernel Trick $\rightarrow$ KRR

Applying the kernel trick using the kernel  $k(.,.)$  we get

$$X X^T \rightarrow \mathbf{K} \quad \text{where} \quad \mathbf{K}_{ij} = k(\vec{x}_i, \vec{x}_j)$$

and

$$X x' \rightarrow \vec{\kappa} \quad \text{where} \quad \kappa_i = k(\vec{x}_i, \vec{x}')$$

**Result:** To use *Kernel Ridge Regression* for training and prediction we need the two results.

#### 1. Training:

Calculate  $\vec{\alpha}$  from training data.

$$\vec{\alpha} = (\mathbf{K} + \lambda I)^{-1} \vec{y} \quad (1)$$

#### 2. Prediction:

Predict new value  $y'$  using

$$y' = \vec{\alpha}^T \vec{\kappa} \quad (2)$$

### Kernel function

One choice is the *Gaussian kernel*

$$k(\vec{x}, \vec{x}') = \exp \left( \frac{d(\vec{x}, \vec{x}')^2}{2\sigma^2} \right) \quad (3)$$

where

$$d(\vec{x}, \vec{x}') = \sqrt{\sum_k^M (x_k - x'_k)^2} \quad (4)$$

### KRR prediction-error estimate

The estimate of the prediction error on a new point  $x'$  is

$$\text{err}(x') = \sqrt{|\theta_0 [1 - \vec{\kappa}(x') \cdot \vec{\alpha}_2(x')]|} \quad (5)$$

where

$$\vec{\alpha}_2(x') = (\mathbf{K} + \lambda I)^{-1} \vec{\kappa}(x')$$

and

$$\theta_0 = \frac{y \cdot \alpha_1}{N_{train}}$$

## 2 Feature