

CS/ECE/ME532 Period 24 Activity

Estimated Time: 30 minutes for P1, 30 minutes for P2, 10 minutes for P3

- 1. Kernel regression.** Kernel regression predicts a value d corresponding to value x as $\hat{d}(x) = \sum_{i=1}^N \alpha_i K(x, x^i)$ where the measured data is $(d^i, x^i), i = 1, 2, \dots, N$ and $K(u, v)$ is the kernel function. We will assume Gaussian kernels, $K(u, v) = \exp(-(u - v)^2/(2\sigma^2))$. Scripts are provided to help you explore properties of kernel regression with respect to the kernel parameter σ and ridge regression parameter λ .

- a) Run the regression script with $\sigma = 0.04$ and $\lambda = 0.01$. Figure 1 displays several of the kernels $K(x, x^i)$. What is the value x^i associated with the kernel having the third peak from the left? What property of the kernel is determined by x^i ? What property is determined by σ ?
- b) Run the regression script for the following choices of regularization and kernel parameters:
 - i. $\lambda = 0.01, \sigma = 0.04$
 - ii. $\lambda = 0.01, \sigma = 0.2$
 - iii. $\lambda = 0.01, \sigma = 1$
 - iv. $\lambda = 1, \sigma = 0.04$
 - v. $\lambda = 1, \sigma = 0.2$

(Note that you need to rerun the entire script each time to ensure the random number generator is reset and you obtain identical data.) You may choose additional cases if it helps you understand the nature of the solution. Discuss how λ and σ affect the characteristics of the kernel regression to the measured data, and support your conclusions with rationale and plots.

- c) What principle could you apply to select appropriate values for λ and σ ?

- 2. Kernel Classification.** The kernel classification script performs classification using the squared error loss using the Gaussian kernel $K(\mathbf{u}, \mathbf{v}) = \exp(-\|\mathbf{u} - \mathbf{v}\|_2^2/(2\sigma^2))$ for the two classification problems we studied using the three-layer neural network in Period 22 Activity. The code is set up to use $N=1000$ training samples, which means the classifier requires a 1000 by 1000 matrix inverse. If this runs too slow on your computer, you may change N to 500.

Run the code for the following values of the kernel parameter σ .

- a) $\sigma = 5$
- b) $\sigma = 0.05$

c) $\sigma = 0.005$

Use the results to discuss the impact of the kernel parameter σ on the decision boundary. Is there a downside to choosing a very small value for σ ? *Hint:* Consider the amplitudes of the predicted labels as σ decreases. Run additional values for σ if needed.

3. **SVM.** You use a kernel-based support vector machine for binary classification with labels $d^i = \{+1, -1\}$. Given training features and labels $(\mathbf{x}^i, d^i), i = 1, 2, \dots, N$ you use a kernel $K(\mathbf{u}, \mathbf{v})$ and design the classifier weights $\boldsymbol{\alpha}$ as

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \sum_{i=1}^N \left(1 - d^i \sum_{j=1}^N \alpha_j K(\mathbf{x}^i, \mathbf{x}^j) \right)_+ + \lambda \sum_{i=1}^N \sum_{j=1}^n \alpha_i \alpha_j K(\mathbf{x}^i, \mathbf{x}^j)$$

- a) Assume the optimization problem has been solved to obtain the weights $\boldsymbol{\alpha}$. Express the classification procedure for a measured feature \mathbf{x} .
- b) Suppose $N = 1000$ and $\alpha_i = 0, i = 1, 2, \dots, 99, 102, 103, \dots, 1000$. Identify the support vectors and write the classification procedure in terms of the support vectors.