CS/ECE/ME 532

Unit 6 Practice Problems

1. Neural Network Basic. Consider the following neural network.

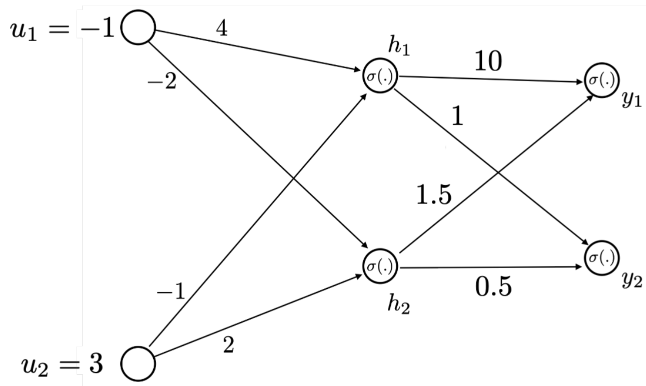

Figure 1: A Neural Network Architecture

a) How many hidden layers are in this network? How many neurons are in this network? How many outputs does this network have? What is $\sigma(\cdot)$ called?

b) Assume that the neural network shown above uses an activation called *Leaky ReLU*, defined as follows:

$$\sigma(z) = \begin{cases} z & \text{if } z > 0 \\ 0.01z & \text{otherwise.} \end{cases}$$

Note that Leaky ReLU behaves differently from ReLU when the input value is less than 0. The values labeled h_1, h_2, y_1, y_2 are the outputs of the corresponding nodes. Find the numerical values of h_1, h_2, y_1, y_2 for an input feature $u_1 = -1, u_2 = 3$.

2. SGD to learn the weights for a single neuron We use the single neuron shown in the figure for classification. Here x_j^i is the j-th feature in the i-th training sample and the output is $\widehat{y}^i = \sigma\left(\sum_{j=0}^{P} w_j x_j^i\right)$, where the activation function is *ReLU6*, defined as follows:

$$\sigma(z) = \min(\max(0, z), 6).$$

Note that ReLU6 behaves differently from ReLU when the input value is larger than 6.

Now suppose we use ridge regression for the loss function

$$f^i(\boldsymbol{w}) = \frac{1}{2}(\widehat{y}^i - y^i)^2 + \lambda \sum_{j=0}^{P} w_j^2.$$

Derive the gradient for the update step, $\nabla f^{i_t}(\boldsymbol{w}^{(t)})$ and write the update equation for $\boldsymbol{w}^{(t+1)}$.

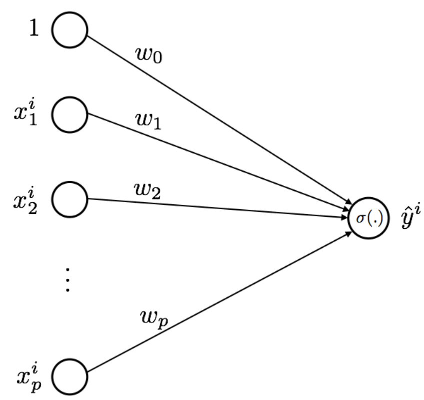

Figure 2: A Neuron

3. Kernel Regression. Consider the Gaussian kernel regression with ℓ_2 regulariztaion. The first hyperparameter is σ for the Gaussian kernel $K(\boldsymbol{u}, \boldsymbol{v}) = \exp\left(-||\boldsymbol{u} - \boldsymbol{v}||_2^2/(2\sigma^2)\right)$. The second hyperparameter is λ which is the weight for the ℓ_2 regularization term.

Now, consider the following five choices:

a) $\lambda = 0.01, \sigma = 0.04$

b) $\lambda = 0.01, \sigma = 0.2$

c) $\lambda = 0.01, \sigma = 1$

d) $\lambda = 1, \sigma = 0.04$

e) $\lambda = 1, \sigma = 0.2$

Find the correct mapping between each of these choices and each of the following options (each row corresponds one choice).

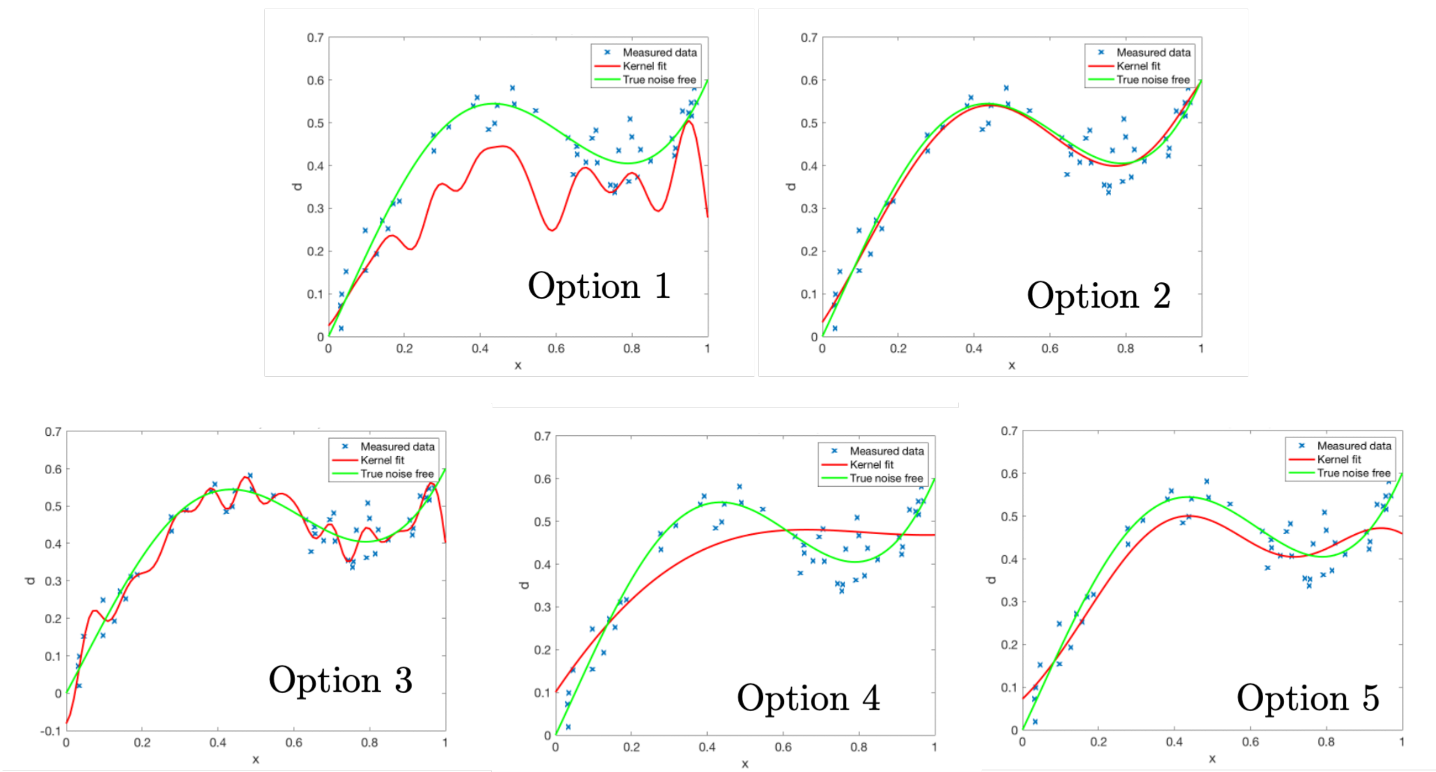

Figure 3: Kernel regression results with five different hyperparameters

4. Kernel Classification. Consider the kernel classification with the squared error loss using the Gaussian kernel $K(\boldsymbol{u}, \boldsymbol{v}) = \exp\left(-||\boldsymbol{u} - \boldsymbol{v}||_2^2/(2\sigma^2)\right)$ for the following dataset.

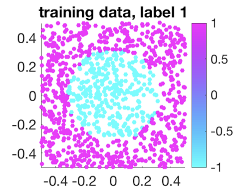

Figure 4: Binary classification dataset

Now, consider the following three choices:

a) $\sigma = 5$

b) $\sigma = 0.05$

c) $\sigma = 0.005$

Find the correct mapping between each of these choices and each of the following options (each row corresponds one choice).

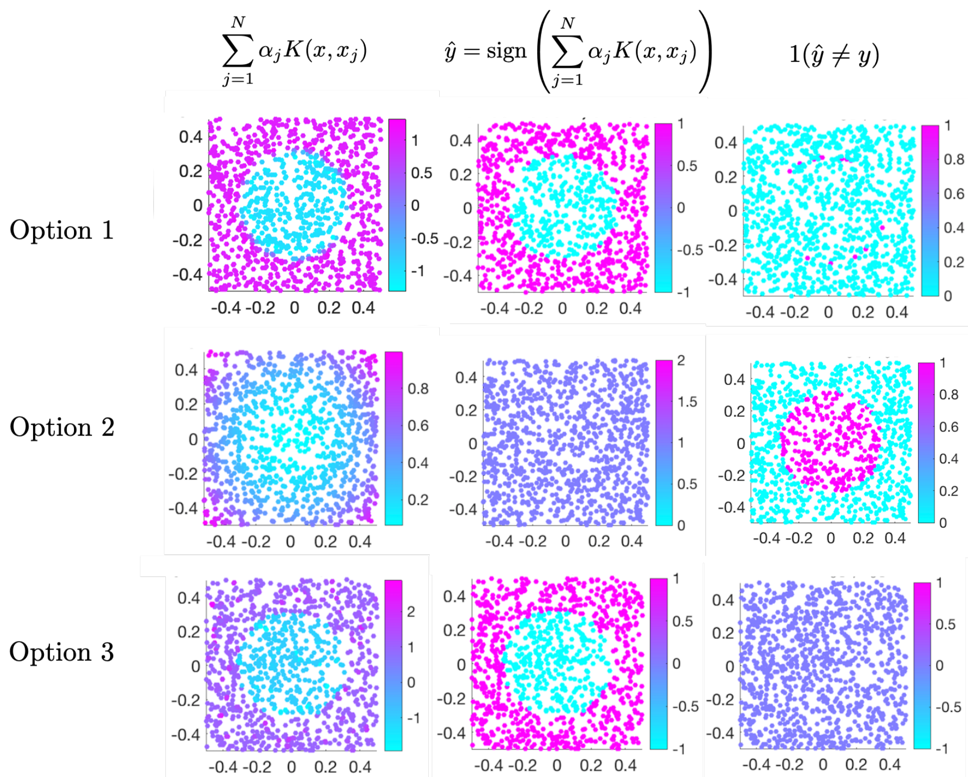

Figure 5: Kernel classification results with three different options

5. **Kernel Classification.** When using a Gaussian kernel, is there a downside to choosing a very small value for σ?

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

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

a) Assume the optimization problem has been solved to obtain the weights $\boldsymbol{\alpha}$. Express the classification procedure for a measured feature $\boldsymbol{x}$.

b) Suppose $N = 1000$ and $\alpha_i = 0, i = 1, 2, \ldots, 99, 102, 103, \ldots, 1000$. Identify the support vectors and write the classification procedure in terms of the support vectors.