Exercises
Learning and Intelligent Systems
SS 2016

Series 2, Mar 15th, 2016 (Kernels)

LAS Group, Institute for Machine Learning
Dept. of Computer Science, ETH Zürich
Prof. Dr. Andreas Krause
Web: http://las.inf.ethz.ch/teaching/lis-s16/
Email questions to:
Adish Singla, adish.singla@inf.ethz.ch

It is not mandatory to submit solutions and sample solutions will be published in two weeks. If you choose to submit your solution, please send an e-mail from your `ethz.ch` **address with subject** `Exercise2` **containing a PDF (LaTeX or scan) to** `lis2016@lists.inf.ethz.ch` **until Sunday, April 3rd, 2016.**

Problem 1 (Kernel Composition):

Assume that $k_i : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $i = 1, 2$, are kernels with corresponding features mappings $\Phi_i : \mathcal{X} \to \mathcal{F}_i$ to some features spaces $\mathcal{F}_i$. For each definition of $k(\cdot, \cdot)$ below, prove that k is also a kernel by finding the corresponding mapping $\Phi : \mathcal{X} \to \mathcal{F}$ to a feature space $\mathcal{F}$.

(a) $k(\boldsymbol{x}, \boldsymbol{y}) := ak_1(\boldsymbol{x}, \boldsymbol{y})$, for some $a > 0$.

(b) $k(\boldsymbol{x}, \boldsymbol{y}) := k_1(\boldsymbol{x}, \boldsymbol{y}) + k_2(\boldsymbol{x}, \boldsymbol{y})$.

(c) $k(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^T \mathbf{M} \boldsymbol{y}$, for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, and some symmetric positive semidefinite $\mathbf{M} \in \mathbb{R}^{d \times d}$.

Problem 2 (Kernelized Linear Regression):

In this exercise you will derive the kernelized version of linear regression.

(a) Prove that the following identity holds for any matrix $\mathbf{B} \in \mathbb{R}^{n \times m}$, and any invertible matrices $\mathbf{A} \in \mathbb{R}^{m \times m}$, and $\mathbf{C} \in \mathbb{R}^{n \times n}$.

$$\left(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B}\right)^{-1} \mathbf{B}^T \mathbf{C}^{-1} = \mathbf{A} \mathbf{B}^T \left(\mathbf{B} \mathbf{A} \mathbf{B}^T + \mathbf{C}\right)^{-1}$$

(b) Remember the solution of ridge regression, $\boldsymbol{w}^* = \left(\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I}\right)^{-1} \mathbf{X}^T \boldsymbol{y}$. Use the matrix identity of part (a) to prove that $\boldsymbol{w}^*$ lies in the row space of $\mathbf{X}$, that is, it can be written as $\boldsymbol{w}^* = \mathbf{X}^T \boldsymbol{z}^*$ for some $\boldsymbol{z}^* \in \mathbb{R}^n$.

(c) Use the result of part (b) to transform the original ridge regression loss function,

$$R(\boldsymbol{w}) = \|\mathbf{X}\boldsymbol{w} - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{w}\|_2^2,$$

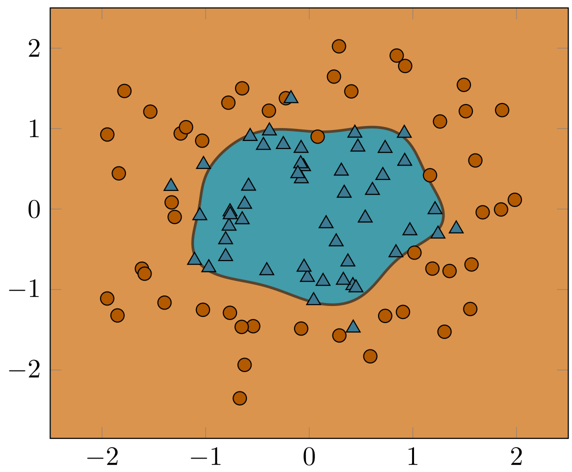
into a new loss function $\hat{R}(\boldsymbol{z})$, such that $\hat{R}(\boldsymbol{z}^*) = R(\boldsymbol{w}^*)$, and $\boldsymbol{z}^* = \operatorname{argmin}_{\boldsymbol{z}} \hat{R}(\boldsymbol{z})$.

(d) Assuming that you are given a kernel $k(\cdot, \cdot)$, express the kernel matrix $\mathbf{K}$ of the data set as a function of the data matrix $\mathbf{X}$, and substitute it in the new loss function $\hat{R}(\boldsymbol{z})$ to obtain the kernelized version of the ridge regression loss function.

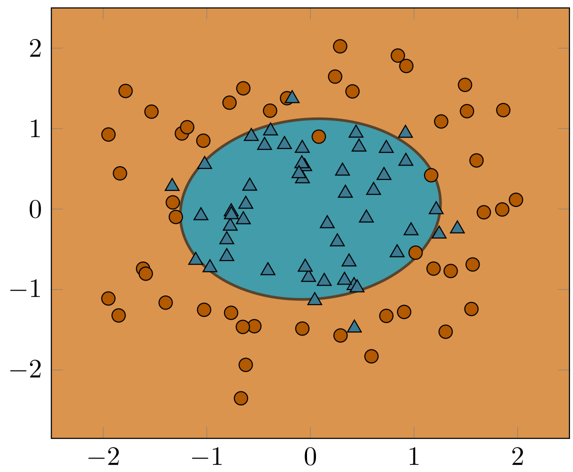
(e) To complete the kernelized version of ridge regression, show how you would predict the value y of a new point $\boldsymbol{x}$, assuming that you have already computed $\boldsymbol{z}^*$.

Problem 3 (Classifiers):

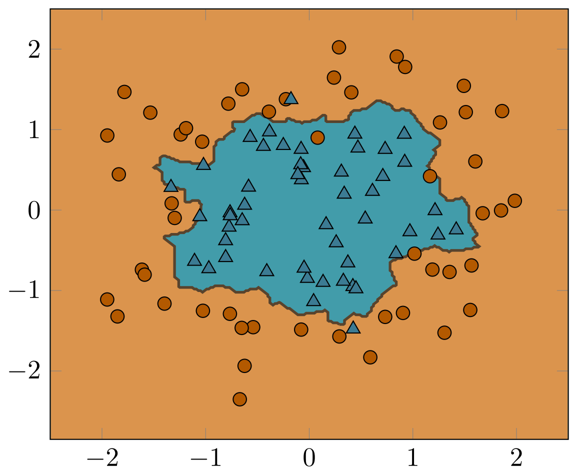
The following figure shows three classifiers trained on the same data set. One of them is a k-nearest neighbor classifier, and the other two are support vector machines (SVMs) using a quadratic and a Gaussian kernel respectively. Based on the shape of the decision boundary, can you guess which plot corresponds to which classifier?



(a)



(b)



(c)