Boston University

ENG EC 414 Introduction to Machine Learning

Exam 2

Released on Wednesday, 11 November, 2020 (120 minutes, 41 points + 2 bonus points)

Submit to Gradescope

- *There are 6 problems plus 1 bonus one.*
- *Unless explicitly written, for each part you must clearly outline the key steps and provide proper justification for your calculations in order to receive full credit.*
- *You can use any material from the class (slides, discussions, homework solutions, etc.), but you cannot look for solutions on the internet. Also, be aware of the limited time.*

Problem 2.1 [10pts] Consider the following 6 training feature vectors: $\mathbf{x}_A = (2, 0)^\top, \mathbf{x}_B = (3, 2)^\top, \mathbf{x}_C = (4, 1)^\top, \mathbf{x}_D = (1, 2)^\top, \mathbf{x}_E = (2, 4)^\top, \mathbf{x}_F = (0, 3)^\top$ with *class* labels $+1, +1, +1, -1, -1, -1$ respectively.

(a) [2pts] Hand-plot the training set and properly label axes and key points.

(b) [3pts] Compute the coordinates of the point in the convex hull of negative training feature vectors which is closest (in Euclidean distance) to the point $\mathbf{x}_B$.

(c) [3pts] Hand-compute the parameters of the hard-margin SVM hyperplane for this training set in *canonical* form. Sketch the SVM hyperplane.

(d) [2pts] Compute the size of the margin of the hard-margin SVM.

Problem 2.2 [4pts] (True/False) For each statement, say if it is true or false. No justification is necessary here.

(i) [2pts] (SVM)

(a) The objective function of hard-margin SVM is convex.

(b) The hyperplane of an SVM $\mathbf{w}_{SVM}$ is a linear combination of training samples in both hard- and soft-margin SVMs.

(c) If we run a soft-margin SVM on a linearly separable dataset, we always get training error zero.

(d) SVMs cannot be kernelized due to the nonlinear hinge loss term.

(ii) [2pts] In polynomial regression with squared loss, as the degree increases, the *training* error:

(a) is non-decreasing

(b) is non-increasing

(c) increases first and then decreases

(d) decreases first and then increases

Problem 2.3 [4pts] Say if each statement is true or false and **explain your choices to get full credit**.

Let $K(\mathbf{u}, \mathbf{v})$ a kernel, then:

(a) K is symmetric, that is $K(\mathbf{u}, \mathbf{v}) = K(\mathbf{v}, \mathbf{u}), \forall \mathbf{u}, \mathbf{v}$

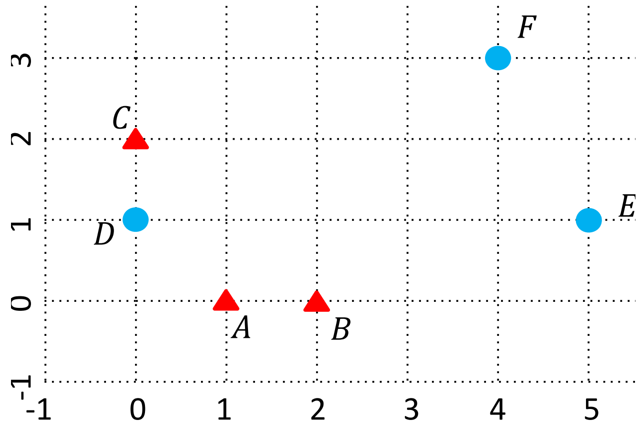
(b) K is non-negative, that is $K(\mathbf{u}, \mathbf{v}) \geq 0, \ \forall \mathbf{u}, \mathbf{v}$

(c) There exists a unique function ϕ such that $K(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{u})^\top \phi(\mathbf{v})$

(d) $K(\mathbf{u}, \mathbf{u}) \geq 0, \ \forall \mathbf{u}$.

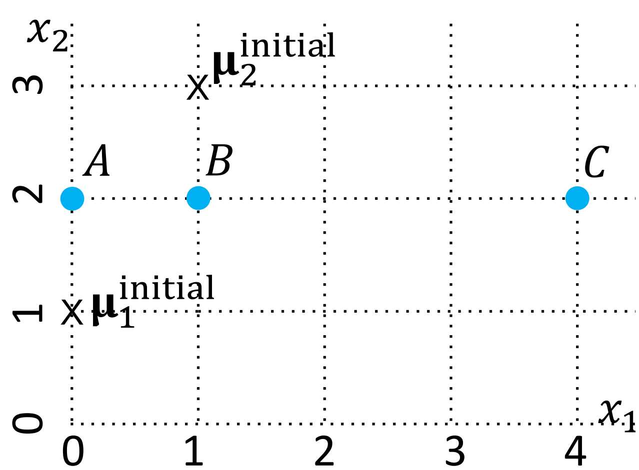
Problem 2.4 [6pts] Let $\mathbf{x}_A = (2, 0)^\top$, $\mathbf{x}_B = (-2, 0)^\top$, and $\mathbf{x}_C = (0, 2)^\top$ be three points in $\mathbb{R}^2$. Compute and sketch the Voronoi-tessellation of $\mathbb{R}^2$ induced by the three points for **Euclidean** distance.

Problem 2.5 [7pts] A training set with points A, B, C in one class and D, E, F in another is shown in



the figure. To select a value of k for k-NN classification using **Euclidean** distance, we perform leave one out cross-validation (LOOCV). (i) For each training point, list all its 5 nearest neighbors in the order of increasing Euclidean distance. (ii) For $k = 1, 3, 5$, list the validation examples that will be misclassified and the corresponding LOOCV error. Identify the best value of k among these choices.

Problem 2.6 [10pts] For the dataset of 3 points A, B, C and initial mean vectors $\mu_1^{\text{initial}}, \mu_2^{\text{initial}}$ shown



in the figure, compute the clusters and mean vectors found by running the 2-means algorithm (Euclidean distance) until convergence. In particular, complete the following table:

Iteration number	**Clusters**	$\boldsymbol{\mu}_1$	$\boldsymbol{\mu}_2$
(Initialization)		$(0, 1)^\top$	$(1, 3)^\top$
...	...	...	...

Problem 2.7 [Bonus, 2pts] Consider x, y integers where $x, y \leq 100$. Show that $K(x, y) = \min(x, y)$ is a valid kernel finding the corresponding transformation ϕ. Hint: Consider $\phi : \mathbb{R} \rightarrow \mathbb{R}^{100}$.