Midterm
Group A

Problem 1 (20 points)

Recall D-dimensional regression problem when your model performs label

prediction for the i-th example x_i in the training data set using linear function:

$$f(x_i; \theta_0, \theta_1, \theta_2, \ldots, \theta_D) = \sum_{d=1}^{D} \theta_d x_i(d) + \theta_0.$$

In this case x_i is D-dimensional. Let y_i denote the true label of the i-th example and let N be the total number of training examples. Parameters of the model $(\theta_0, \theta_1, \theta_2, \ldots, \theta_D)$ are obtained by minimizing the empirical risk provided below:

$$R(\theta) = \frac{1}{2N} ||y - X\theta||_2^2 + \theta^T H \theta + \theta^T \theta + a^T \theta,$$

where a is a vector and H is a matrix that satisfies the condition: $H = H^T$. Both a and H are given. Write what is y, X, and θ in the formula above. Compute the optimal setting of parameters by setting the gradient of the risk to 0. Explain all steps in your derivations.

Problem 2 (15 points)

A kernel is an efficient way to write out an inner product between two feature vectors computed from a pair of input vectors as follows:

$$K(x, y) = \phi(x)^\top \phi(y).$$

Assume that both inputs are 2-dimensional and write out the explicit mapping ϕ that mimics the kernel value for a 3^{rd}-order polynomial kernel as follows:

$$K(x, y) = (x^\top y + 1)^3.$$

Problem 3 (15 points)

The exponential distribution has density given as

$$p_\lambda(x) = \lambda e^{-\lambda x}$$

Find the maximum likelihood estimator for λ. Calculate an estimate using this estimator when $x_1 = 1,\ x_2 = 2,\ x_3 = 4,\ x_4 = 2.$

Problem 4 (15 points)

Consider 2d family of classifiers given by an origin-centered circles $f(x) = sign(ax^\top x + b)$. What is the VC dimension of this family? Prove it.

Problem 5 (15 points)

Using the principle of Lagrange multipliers, find the maximum and minimum values of $f(x, y) = x^2 - y^2$ subject to the constraint, $x^2 + y^2 = 1$.

Problem 6 (10 points)

What is the maximum likelihood estimator of the mean μ and covariance matrix Σ of the following 2-dimensional data set $\mathcal{X}$? Justify your answer.

$$\mathcal{X} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 7 \\ 8 \end{bmatrix}, \begin{bmatrix} 9 \\ 10 \end{bmatrix} \right\}$$

Problem 7 (10 points)

Explain the difference between overfitting and underfitting. For a picture given below, show and example of underfitting on the left plot, proper fit in the middle plot, and overfitting on the right plot.