Midterm
Group B

Problem 1 (20 points)

Consider 1-dimensional polynomial regression problem when your model

performs label prediction for the i-th example x_i in the training data set using polynomial function:

$$f(x_i; \theta_0, \theta_1, \theta_2, \ldots, \theta_P) = \sum_{p=1}^{P} \theta_p x_i^p + \theta_0.$$

In this case x_i is 1-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 $(\theta_0, \theta_1, \theta_2, \ldots, \theta_P)$ are obtained by minimizing the empirical risk given below:

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

for some vector a that is 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)

Suppose that $w = xy^2$ and $x = r\cos\theta$, $y = r\sin\theta$.

Use the chain rule to find $\partial w/\partial r$ when $r = 2$, $\theta = \pi/4$.

Problem 3 (15 points)

Suppose $X_1, X_2, \ldots, X_n$ are i.i.d. samples from a population with pdf

$$f(x|\theta) = \frac{1}{\theta} x^{(1-\theta)/\theta}, \quad 0 < x < 1, \quad 0 < \theta < \infty.$$

Find the maximum likelihood estimator for θ. Calculate an estimate using this estimator when $x_1 = 0.10$, $x_2 = 0.22$, $x_3 = 0.54$, $x_4 = 0.36$.

Problem 4 (15 points)

Find the extreme values of the function $f(x, y) = xe^y$ subject to the constraint $x^2 + y^2 = 2$. Use Lagrange multipliers.

Problem 5 (15 points)

Consider 2d family of classifiers given by axis-aligned rectangles. What is the VC dimension of this family?

Problem 6 (10 points)

Suppose we have m boxes. i^{th} box ($i = \{1, 2, \ldots, m\}$) contains n_i apples and r_i oranges. One of the boxes is chosen at random (with equal probability of choosing any box) and an item is selected from the box and found to be an apple. Use Bayes' rule to find the probability that the apple came from the j^{th} box ($j \in \{1, 2, \ldots, m\}$).

Problem 7 (10 points)

Consider the following plot, where we fit the polynomial of order M ($f(x; w) = \sum_{j=0}^{M} w_j x^j$) to the dataset, where $w = [w_0 \;\; w_1 \;\; \ldots \;\; w_M]^\top$ denotes the vector of model weights and the dataset is a collection of 2-dimensional points (x, y). The dataset is represented with the blue circles on the figure.

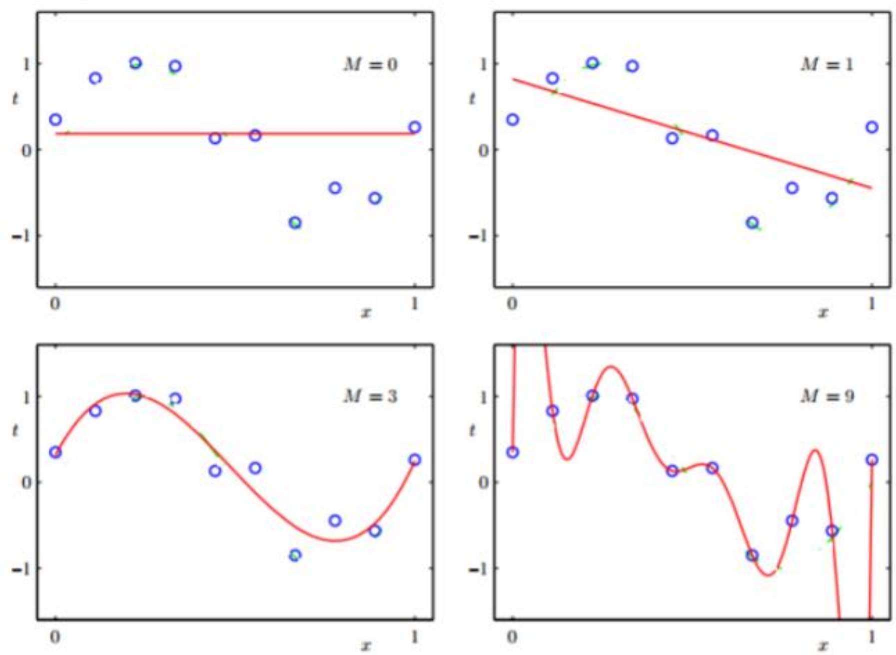

Figure 1: Plots of polynomials having various orders M, shown as red curves, fitted to the data set.

What is the reasonable choice of M and why? Which M correspond to overfitting (7 points) and which to underfitting and why?

Consider any loss function that measures the discrepancy between the target values and the predictions of the model, e.g. squared loss which for a single data point is defined as $L(y_i, f(x_i, w)) = \frac{1}{2}(y_i - f(x_i, w))^2$). Draw a typical behavior of the train and test loss for the optimal setting of model weights as a function of M, where recall that the train loss is the loss computed for a training dataset (the model was trained on this dataset) and the test loss is the loss computed for a test dataset (the model did not see this dataset during training). Indicate overfitting and underfitting regimes.