In problems 5, 6, 8, and 9 for full credit verbally describe the steps of your derivation.

Problem 1 (5 points)

List at least two loss functions and describe the context in which they will be most useful.

Problem 2 (5 points)

Describe 5 features that can be used to decide whether a fruit is an apple or an orange.

Problem 3 (10 points)

Vector $(3, -7, 5)$ is normal to a plane P and point $(0, 0, -2)$ is in the plane. Provide equation of P.

Problem 4 (10 points)

Establish that for any n-dimensional vector $\boldsymbol{x}$ its length denoted by $|\boldsymbol{x}|$ equals to $\sqrt{\boldsymbol{x}^T\boldsymbol{x}}$. Please, do not give recursive definition of the type: length is an L_2 norm therefore it is length. Think from the basic principles. *Start baby steps, what is length in the real world? How does it generalize to higher dimension? Does the basis in which $\boldsymbol{x}$ is defined matter?*

Problem 5 (15 points)

For a fixed classifier h in the class of binary classifiers $\mathcal{H}$ that operate on domain $\mathcal{X}$ generated according to an unknown distribution $\mathcal{D}$ and labeled by f, show that the expected value of $L_S(h)$ over the choice of training sequences $S\mid_x$ equals to $L_{(\mathcal{D},f)}(h)$, specifically:

$$\mathop{\mathbb{E}}_{S|_x \sim \mathcal{D}^m} [L_S(h)] = L_{(\mathcal{D},f)}(h)$$

Problem 6 (25 points)

Let $\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \{0, 1\}$, and let $\mathcal{H}$ be the class of origin centered rectangles in the plane, that is $\mathcal{H} = \{h_{(a,b)} : a, b \in \mathbb{R}_+\}$, where $h_{(a,b)}(\boldsymbol{x}) = \mathbb{1}_{[|x_1| \leq a, |x_2| \leq b]}$ ($\mathbb{1}$ is an indicator function that returns 1 if $\boldsymbol{x}$ falls inside an origin-centered rectangle with width equal to $2a$ and height equal to $2b$, x_i are components of $\boldsymbol{x}$). Prove that $\mathcal{H}$ is PAC learnable (assume realizability) and show that the sample complexity is bounded by

$$m_{\mathcal{H}}(\epsilon, \delta) \leq \left\lceil \frac{\log 1/\delta}{\epsilon} \right\rceil.$$

Problem 7 (30 points)

Implement the batch and stochastic (single-sample) Perceptron algorithms (provide the code with the solution) and explore the following:

(a) (4 points) Generate 25 3-d points for each of the two classes from Gaussian distributions, $p(\boldsymbol{x}|w_i) \sim N(\mu_i, \mathbf{I})$, and $\boldsymbol{\mu}_1 = \mathbf{0}$, $\boldsymbol{\mu}_2 = (4, 4, 4)^T$

(b) (4 points) Compute $\beta^2 = \max_i \|\boldsymbol{x}_i\|^2$ (3 points)

(c) (5 points) Find a separating hyperplane using the batch algorithm. Describe your observations.

(d) (4 points) Randomly choose an initial weight vector $\boldsymbol{w}_0$ constrained to the surface of a 3D sphere of radius 0.1 (e.g. $\|\boldsymbol{w}_0\| = 0.1$).

(e) (5 points) Explore stochastic perceptron algorithm performance on the obtained data with the starting value of $\boldsymbol{w}_0$ as above. Store the solution $\widetilde{\boldsymbol{w}}$.

(f) (4 points) Compute $\gamma = \min_i [\boldsymbol{w}^T \boldsymbol{x}_i l_i]$, $\alpha = \beta^2/\gamma$, and finally:

$$k_0 = \frac{\|\boldsymbol{w}_0 - \alpha \widetilde{\boldsymbol{w}}\|^2}{\beta^2}$$

(g) (4 points) Compare k_0 with the number of iterations your algorithm took. Experiment with different initialization strategies for $\boldsymbol{w}$ and repeat the experiment.

Problem 8 (extra credit 15 points)

For difficult decisions B. Busyness, a CEO of a Fortune 500 company, relies on expert consultants from the ABCD consulting group. ABCD offers an appropriate consultant out of its $\mathcal{H} = 2800$ experts.

(a) (10 points) Assume that difficult decisions need to be made in a random fashion independently and identically according to some distribution $\mathcal{D}$. Assume that ABCD's HR department can find and select an expert consultant out of $\mathcal{H}$ who has made a decision that left the company they were consulting for happy for the last $m = 200$ times ABCD was hired for the job. Give a bound on the chances that this consultant suggests a decision that will leave B. Busyness unhappy as a result. What is the value of the bound with 95% confidence?

(b) (10 points) Assume now that the HR department can find a consultant out of $\mathcal{H}$ who made profitable advice for all but $m' = 20$ of the last $m = 200$ times they were hired. What is the value of the new bound?

Problem 9 (extra credit 15 points)

The *convex hull* of a set of vectors $\boldsymbol{x}_i, i = 1, \ldots, n$ is the set of all vectors of the form

$$\boldsymbol{x} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}_i,$$

where $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Given two sets of vectors, show that either they are linearly separable or their convex hulls intersect. (To answer this, suppose that both statements are true, and consider the classification of a point in the intersection of the convex hulls.)