Lecture 1

This document provides practice problems that are similar to those that will be asked during the final exam. Please note that the document reflects the style and not the number of the questions that will be on the exam.

Problem 1

Let $f(x;\alpha)$ be a machine learning classifier parametrized by α in the hypothesis class Λ. Let the true risk

$$R(\alpha) = \mathbb{E}_{(X,Y)\sim P}[l(f(X;\alpha), Y)]$$

be empirically approximated by $\hat{R}(\alpha) = \sum_{i=1}^{n} l(f(x_i;\alpha), y_i)$ based on n samples $\{(x_i, y_i)\}_{i=1}^{n}$, which are (mutually) independent and identically distributed. For the 0−1 loss and a finite hypothesis class of size m, we derived the bound

$$R(\alpha) \leq \hat{R}(\alpha) + \sqrt{\frac{\log m - \log \delta}{2n}},$$

with probability at least $1-\delta$ for arbitrary $0 < \delta < 1$. We call the term $\sqrt{(\log m - \log \delta)/(2n)}$ the generalization gap.

Suppose now our hypothesis class consists of the functions $f(x;\alpha) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3$, where $\alpha_i \in \{0, 1, \dots, 5\}$, and let us compare the following two scenarios:

- Scenario 1: We set the bias $\alpha_0 = 0$ and use a training dataset of size $n > 0$.
- Scenario 2: We use a training dataset of size $9n$.

If G_1 and G_2 denote the generalization gap in Scenario 1 and Scenario 2, respectively, then the ratio G_2/G_1 is:

(a) $G_2/G_1 = 1/\sqrt{6}$ for all $0 < \delta < 1$. **[False]**

(b) $G_2/G_1 = \sqrt{6}$ for $\delta = 1/5$. **[False]**

(c) $G_2/G_1 = 2/(3\sqrt{3})$ for all $0 < \delta < 1$. **[False]**

(d) $G_2/G_1 = 3\sqrt{3}/2$ for $\delta = 1/10$. **[False]**

(e) $G_2/G_1 = \sqrt{5}/6$ for $\delta = 1/6$. **[True]**

Explanation: In scenario 1, we are looking at $m_1 = 6^3$ hypotheses (6 options each for $\alpha_1, \alpha_2, \alpha_3$); in scenario 2, we have $m_2 = 6^4$. Plugging this into the formula for the generalization gap (together with $\delta = 1/6 = 6^{-1}$ and the data set sizes) and using calculation rules for logarithms gives $G_2/G_1 = \sqrt{5}/6$.

Problem 2

Let X be a continuous random variable, uniformly distributed in $[1, 2]$. Which of the following statements is true?

(a) $\mathbb{E}[e^X] < e^{\mathbb{E}[X]}$ **[False]**

(b) $\mathbb{E}[e^X] = e^{E[X]}$ **[False]**

(c) $\mathbb{E}[e^X] > e^{\mathbb{E}[X]}$ **[True]**

Explanation: This is a direct application of Jensen's inequality: $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$ for convex φ. For an alternative derivation, consider this:

$$\mathbb{E}[\exp\{X\}] = \mathbb{E}[\exp\{X + \mathbb{E}[X] - \mathbb{E}[X]\}] = \exp\{\mathbb{E}[X]\} \cdot \mathbb{E}[\exp\{X - \mathbb{E}[X]\}] \geq \exp\{\mathbb{E}[X]\} \cdot \mathbb{E}[1 + X - \mathbb{E}[X]] = \exp\{\mathbb{E}[X]\},$$

using the linearity of the expectation, the fact that $\mathbb{E}[c] = c$ for constant c, and the Taylor expansion $\exp(x) = 1 + x + \dots$ to get $\exp(x) \geq 1 + x$.

Note: This result has some important applications in statistical physics; for example, in *mean field theory*, which, in turn, has found applications in machine learning.

Problem 3

Tom plays roulette in a casino. At each round he bets 1 dollar on red. Let $X_i \in \{0, 1\}$ denote the random variable that is equal to 1 if he wins and 0 if he loses at round i, where $i = 1, 2, \ldots, n$. Which of the following statements are true?

(a) The random variables $\{X_i\}_{i=1}^n$ are dependent and identically distributed. **[False]**

(b) The random variables $\{X_i\}_{i=1}^n$ are dependent and not identically distributed. **[False]**

(c) The random variables $\{X_i\}_{i=1}^n$ are independent and identically distributed. **[True]**

Explanation: This follows directly from the fact that (by definition) all rounds are independent.

(d) The random variables $\{X_i\}_{i=1}^n$ are independent and not identically distributed. **[False]**

Note: In roulette, the ball lands uniformly on any of the numbers between 1 and 36 inclusive (we ignore the 0 here). Half of the numbers are black, the others are red. The outcome at a given round is independent of all other rounds.

Problem 4

Tom plays roulette in a casino. At each round he bets 1 dollar either on red or on black. He starts by betting on red. Whenever he loses he changes color (that is, if he lost when playing red, he will play black in the next round and vice versa). Let $X_i \in \{0, 1\}$ denote the random variable that is equal to 1 if he wins and 0 if he loses at round i, where $i = 1, 2, \ldots, n$. Which of the following statements are true?

(a) The random variables $\{X_i\}_{i=1}^n$ are dependent and identically distributed. **[False]**

(b) The random variables $\{X_i\}_{i=1}^n$ are dependent and not identically distributed. **[False]**

(c) The random variables $\{X_i\}_{i=1}^n$ are independent and identically distributed. **[True]**

Explanation: This follows directly from the fact that (by definition) all rounds are independent.

(d) The random variables $\{X_i\}_{i=1}^n$ are independent and not identically distributed. **[False]**

Problem 5

Consider a reinforcement learning agent who lives in an ordered state space $\{1, \ldots, K\}$. In state $x_i = m$, the agent chooses action $a \in \{m, m+1, m+2\}$ uniformly at random and this causes its new state to be $x_{i+1} = 1 + (a \bmod K)$. Consider the agent's state trajectory $x_1, \ldots, x_n$ for some starting state x_1. Are the variables $x_{i+1} - x_i$ (for $i = 1, \ldots, n-1$) independent and identically distributed in general?

(a) Independent and identically distributed. **[False]**

(b) Independent, but not identically distributed. **[False]**

(c) Dependent, but identically distributed. **[False]**

(d) Dependent and not identically distributed. **[True]**

Explanation: Consider, for example, what happens when the agent jumps from state K to state 1 at iteration i. In this case, the difference $x_{i+1} - x_i$ is *negative*. Such a jump (negative $x_{i+1} - x_i$) can only happen in cases, where $x_i \in \{K-2, K-1, K\}$, which concludes that the random variables $x_{i+1} - x_i$, $i = 1, 2, \ldots, n-1$, are not identically distributed. Observing a negative $x_{i+1} - x_i$ at iteration i is (in general) informative for the difference $x_{i+2} - x_{i+1}$ at iteration $i+1$ because (for sufficiently large K) you cannot get two negative differences in a row. This implies that $x_{i+1} - x_i$ are also dependent.

Problem 6

Consider a random walk $x_{i+1} = x_i + c_i$, $i = 0, 1, 2, \ldots$, where $c_i \sim \mathcal{N}(0, 1)$ are drawn independently (i.e., the $c_1, c_2, \ldots$ are mutually independent), and where $x_0 = 0$. Is x_{100} independent of x_{99}?

(a) Yes. **[False]**

(b) No. **[True]**

Explanation: We have $x_{100} \sim \mathcal{N}(0, 100)$ (sum of 100 independent normals with unit variance), but $x_{100} \sim \mathcal{N}(x_{99}, 1)$ when conditioning on x_{99}. Hence x_{100} and x_{99} are not independent.

Problem 7

Let X and Y be two discrete random variables that each take values in $\{0, 1, 2\}$. The following table summarizes their joint probability distribution:

	$Y = 0$	$Y = 1$	$Y = 2$
$X = 0$	1/9	1/9	1/9
$X = 1$	2/9	1/9	0
$X = 2$	0	1/9	2/9

Are X and Y independent?

(a) Yes. **[False]**

(b) No. **[True]**

Explanation: The joint probability distribution does not factorize: $\Pr(X = 1, Y = 0) = 2/9 \neq 1/9 = 1/3 \cdot 1/3 = \Pr(X = 1) \cdot \Pr(Y = 0)$.

Problem 8

We roll a perfect die 5 times. Let $X_i \in \{1, 2, \ldots, 6\}$ be the random variable that describes the outcome of the i-th throw. What is the probability that

$$\min_{i \in \{1,2,\ldots,5\}} X_i < 2\,?$$

(a) $1 - (5/6)^5$ **[True]**

Explanation: Consider the complementary probability: We only have $\min_{i \in \{1,2,\ldots,5\}} X_i \geq 2$ if every time we roll the die we get at least a two. The probability for this is $(5/6)^5$, hence we find: $\Pr\big(\min_{i \in \{1,2,\ldots,5\}} X_i < 2\big) = 1 - \Pr\big(\min_{i \in \{1,2,\ldots,5\}} X_i \geq 2\big) = 1 - (5/6)^5$.

(b) $(5/6)^5$ **[False]**

(c) $(1/6)^5$ **[False]**

(d) $1 - (1/6)^5$ **[False]**