
Problem Set 6 — *Due Friday, December 13, before class starts*
For the Exercise Sessions on Nov 29 and Dec 6

Last name	First name	SCIPER Nr	Points

Problem 1: Bounding The Exploration Bias

- (a) Let $X_1, X_2, \dots, X_n \sim \text{i.i.d. } \mathcal{N}(0, 1)$. Let $Y = \arg\max_i X_i$ and $T \in \{1, 2, \dots, n\}$ is such that
- $$P_{T|Y}(t|y) = \begin{cases} p, & t = y \\ \frac{1-p}{n-1}, & t \neq y \end{cases} \quad \text{for some } p \in [0, 1].$$
1. Compute $I(X; T)$ where $X = (X_1, X_2, \dots, X_n)$. (Hint: write $I(X; T) = H(T) - H(T|X)$. What is the marginal distribution of T ?)
- (b) Let $X_1, \dots, X_4 \sim \text{i.i.d. } \mathcal{N}(0, 1)$ and $X_5 \sim \mathcal{N}(0, 4)$. Let Y and T be as in part (a) with $p = 0.3$.
1. Show that $\Pr(Y = 5) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{8\pi}} (1 - Q(x))^4 e^{-x^2/8} dx$ (where $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$), and find a corresponding numerical approximation (using Mathematica, for example).
 2. Using the previous numerical approximation, find the marginal distributions P_Y and P_T .

Problem 2: Dependence and large error events

In the lecture notes we have seen how to bound the expected generalization error using information measures. With this exercise we will work on large error events and provide bounds on the probabilities of such events. The setting is the same: we observe n iid samples $D = (X_1, \dots, X_n)$ (according to some unknown distribution P) and based on this observation we will choose a hypothesis $w \in W$. We also consider the usual definition of empirical and population risk, *i.e.* given a loss function ℓ , some hypothesis w , $L_D(w) = \frac{1}{n} \sum_{i=1}^n \ell(w, X_i)$, and $L_P(w) = \mathbb{E}_P[\ell(w, X)]$. We are interested in controlling the following quantity:

$$\mathbb{P}(|L_P(W) - L_D(W)| > \epsilon). \quad (1)$$

- (a) Suppose that the loss is such that $\ell(w, x) \in \{0, 1\}$ for every $w \in W$ and $x \in \mathcal{X}$. Suppose also that $|\mathcal{W}| < \infty$, *i.e.*, the number of hypotheses is finite.
1. Show that for every **fixed** $w \in W$ $\mathbb{P}(|L_P(w) - L_D(w)| > \epsilon) \leq 2 \exp(-2n\epsilon^2)$;
 2. Show that

$$\mathbb{P}(|L_P(W) - L_D(W)| > \epsilon) \leq |\mathcal{W}| \cdot 2 \exp(-2n\epsilon^2); \quad (2)$$

Hint: denote with $E = \{(d, w) : |L_P(w) - L_d(w)| > \epsilon\}$.

You have that $\mathbb{P}(|L_P(W) - L_D(W)| > \epsilon) = \mathbb{P}(E) = \sum_{(w,d) \in E} P(w, d)$.

(be careful: $\mathbb{P}(|L_P(W) - L_D(W)| > \epsilon | W = w)$ is not necessarily $\leq 2 \exp(-2n\epsilon^2)$. *Why?*)

(b) Now consider the following information measure, given two discrete random variables X, Y :

$$\mathcal{L}(X \rightarrow Y) = \log \sum_y \max_{x: P_X(x) > 0} P_{Y|X}(y|x). \quad (3)$$

This quantity is known in the literature as Maximal Leakage and quantifies the leakage of information between X and Y .

1. Show that if the alphabet of Y (denoted with \mathcal{Y}) is finite then

$$\mathcal{L}(X \rightarrow Y) \leq \log |\mathcal{Y}|,$$

which distributions achieve the bound with equality?

2. It is possible to show that

$$\mathcal{L}(X \rightarrow Y) \geq 0,$$

which distributions achieve the bound with equality?

3. Let X be a binary random variable and let Y be an observation of X after passing through a Binary Symmetric Channel with parameter δ . More precisely we have $P_{Y|X=x}(x) = 1 - \delta$, for $x \in \{0, 1\}$.

What is the maximal leakage $\mathcal{L}(X \rightarrow Y)$?

Which values of δ allow you to achieve the bounds in (1), (2) with equality?

4. Suppose further that the space of samples \mathcal{D} is finite. Denote with $E_w = \{d : (d, w) \in E\}$, for $w \in \mathcal{W}$; Show that:

$$\mathbb{P}(|L_P(W) - L_D(W)| > \epsilon) \leq \exp(\mathcal{L}(D \rightarrow W)) \max_{w \in \mathcal{W}} \mathbb{P}(E_w);$$

5. Conclude that

$$\mathbb{P}(|L_P(W) - L_D(W)| > \epsilon) \leq 2 \exp(\mathcal{L}(D \rightarrow W) - 2n\epsilon^2);$$

6. Compare the two bound retrieved in (a2) and (b4), what do you notice? Is one of the two better than the other? When are they equal? What conclusions can you draw?

Problem 3: MMSE Estimation

Consider the scenario where $p(x|d) = de^{-dx}$, for $x \geq 0$ (and zero otherwise), that is, the observed data x is distributed according to an exponential with mean $1/d$. Moreover, the desired variable d itself is also exponentially distributed, with mean $1/\mu$.

(a) Find the MMSE estimator of d given x , and calculate the corresponding mean-squared error incurred by this estimator.

(b) Find the MAP estimator of d given x .

Problem 4: FIR Wiener Filter

Consider a (discrete-time) signal that satisfies the difference equation $d[n] = 0.5d[n-1] + v[n]$, where $v[n]$ is a sequence of uncorrelated zero-mean unit-variance random variables. We observe $x[n] = d[n] + w[n]$, where $w[n]$ is a sequence of uncorrelated zero-mean random variables with variance 0.5.

(a) (you may skip this at first and do it later — it is conceptually straightforward) Show that for this signal model, the autocorrelation function of the signal $d[n]$ is

$$\mathbb{E}[d[n]d[n+k]] = \frac{4}{3} \left(\frac{1}{2}\right)^{|k|}, \quad (4)$$

and thus the autocorrelation function of the signal $x[n]$ is

$$\mathbb{E}[x[n]x[n+k]] = \begin{cases} \frac{11}{6}, & \text{for } k = 0, \\ \frac{4}{3} \left(\frac{1}{2}\right)^{|k|}, & \text{otherwise.} \end{cases} \quad (5)$$

(b) We would like to find an (approximate) linear predictor $\hat{d}[n+3]$ using only the observations $x[n], x[n-1], x[n-2], \dots, x[n-p]$. Using the Wiener Filter framework, determine the optimal coefficients for the linear predictor. Find the corresponding mean-squared error for your predictor.

(c) We would like to find a linear denoiser $\hat{d}[n]$ using *all* of the samples $\{x[k]\}_{k=-\infty}^{\infty}$. Find the filter coefficients and give a formula for the incurred mean-squared error.