Problem Set 8

For the Exercise Session on Dec 17

Last name	First name	SCIPER Nr	Points

Problem 1: Choose the Shortest Description

Suppose $\mathcal{C}_0 : \mathcal{U} \to \{0,1\}^*$ and $\mathcal{C}_1 : \mathcal{U} \to \{0,1\}^*$ are two prefix-free codes for the alphabet $\mathcal{U}$. Consider the code $\mathcal{C} : \mathcal{U} \to \{0,1\}^*$ defined by

$$\mathcal{C}(u) = \begin{cases} [0, \mathcal{C}_0(u)] & \text{if length}\mathcal{C}_0(u) \leq \text{length}\mathcal{C}_1(u) \\ [1, \mathcal{C}_1(u)] & \text{else.} \end{cases}$$

Observe that $\text{length}(\mathcal{C}(u)) = 1 + \min\{\text{length}(\mathcal{C}_0(u)), \text{length}(\mathcal{C}_1(u))\}$.

(a) Is $\mathcal{C}$ a prefix-free code? Explain.

(b) Suppose $\mathcal{C}_0, \ldots, \mathcal{C}_{K-1}$ are K prefix-free codes for the alphabet $\mathcal{U}$. Show that there is a prefix-free code $\mathcal{C}$ with

$$\text{length}(\mathcal{C}(u)) = \lceil \log_2 K \rceil + \min_{0 \leq k < K-1} \text{length}(\mathcal{C}_k(u)).$$

(c) Suppose we are told that U is a random variable taking values in $\mathcal{U}$, and we are also told that the distribution p of U is one of K distributions $p_0, \ldots, p_{K-1}$, but we do not know which. Using (b) describe how to construct a prefix-free code $\mathcal{C}$ such that

$$\mathbb{E}[\text{length}(\mathcal{C}(U))] \leq \lceil \log_2 K \rceil + 1 + H(U).$$

[Hint: From class we know that for each k there is a prefix-free code $\mathcal{C}_k$ that descibes each letter u with at most $\lceil -\log_2 p_k(u) \rceil$ bits.]

Problem 2: Tighter Generalization Bound

[10pts] Let $D = X_1, ..., X_n$ iid from an unknown distribution P_X, let $\mathcal{H}$ be a hypothesis space, and $\ell : \mathcal{H} \times \mathcal{X} \to \mathbb{R}$ be a σ^2-subgaussian loss function for every h. In the lecture we have seen that the generalization error can be upper bounded using the mutual information.

$$|\mathbb{E}_{P_{DH}}[L_{P_X}(H) - L_D(H)]| \leq \sqrt{\frac{2\sigma^2 I(D;H)}{n}}$$

(i) Modify the proof of the *Mutual Information Bound (11.2.2)* to show that if for all $h \in \mathcal{H}$, $\ell(h, X)$ is σ^2-subgaussian in X, then

$$|\mathbb{E}_{P_{DH}}[L_{P_X}(H) - L_D(H)]| \leq \sqrt{\frac{2\sigma^2 \sum_{i=1}^n I(X_i;H)}{n}}.$$

Hint: Recall from the lecture notes that

$$|\mathbb{E}_{P_{DH}}[L_{P_X}(H) - L_D(H)]| \leq \frac{1}{n} \sum_{i=1}^n \left| \mathbb{E}_{P_{X_i H}}[\ell(H, X_i)] - \mathbb{E}_{P_{X_i} P_H}[\ell(H, X_i)] \right|.$$

(ii) Show that, this new bound is never worse than the previous bound by showing that,

$$I(D; H) \geq \sum_{i=1}^{n} I(X_i; H).$$

(iii) Let us consider an example. Assume that $D = X_1, .., X_n,\ n > 1$, are i.i.d. from $\mathcal{N}(\theta, 1)$, and that we do not know θ. We want to learn θ assuming the loss $\ell(h, x) = \min(1, (h - x)^2)$ (which is bounded) and $\mathcal{H} = \mathbb{R}$. Our learning algorithm outputs $H = \frac{1}{n}\sum_{i=1}^{n} X_i$. Use the new bound to show that

$$|\mathbb{E}_{P_{DH}} [L_{P_X}(H) - L_D(H)]| \leq \sqrt{\frac{1}{4(n-1)}}.$$

How does the old bound perform in this example?
Hint: Adding independent gaussian random variables, you get a gaussian random variable.

Problem 3: Lower bound on Expected Length

Suppose U is a random variable taking values in $\{1, 2, \dots\}$. Set $L = \lfloor \log_2 U \rfloor$. (I.e., $L = j$ if and only if $2^j \leq U < 2^{j+1}$; $j = 0, 1, 2, \dots$.

(a) Show that $H(U|L = j) \leq j$, $j = 0, 1, \dots$.

(b) Show that $H(U|L) \leq \mathbb{E}[L]$.

(c) Show that $H(U) \leq \mathbb{E}[L] + H(L)$.

(d) Suppose that $\Pr(U = 1) \geq \Pr(U = 2) \geq \dots$. Show that $1 \geq i \Pr(U = i)$.

(e) With U as in (d), and using the result of (d), show that $\mathbb{E}[\log_2 U] \leq H(U)$ and conclude that $\mathbb{E}[L] \leq H(U)$.

(f) Suppose that N is a random variable taking values in $\{0, 1, \dots\}$ with distribution p_N and $\mathbb{E}[N] = \mu$. Let G be a geometric random variable with mean μ, i.e., $p_G(n) = \mu^n/(1 + \mu)^{1+n}$, $n \geq 0$.

Show that $H(G) - H(N) = D(p_N \| p_G)$, and conclude that $H(N) \leq g(\mu)$ with $g(x) = (1 + x)\log_2(1 + x) - x \log_2 x$.

[Hint: Let $f(n, \mu) = -\log_2 p_G(n) = (n + 1)\log_2(1 + \mu) - n \log_2(\mu)$. First show that $\mathbb{E}[f(G, \mu)] = \mathbb{E}[f(N, \mu)]$, and consequently $H(G) = \sum_n p_N(n) \log_2(1/p_G(n))$.]

(g) Show that for U as in (d) and $g(x)$ as in (f),

$$E[L] \geq H(U) - g(H(U)).$$

[Hint: combine (f), (e), (c).]

(h) Now suppose U is a random variable taking values on an alphabet $\mathcal{U}$, and $c : \mathcal{U} \to \{0, 1\}^*$ is an injective code. Show that

$$E[\text{length}\, c(U)] \geq H(U) - g(H(U)).$$

[Hint: the best injective code will label $\mathcal{U} = \{a_1, a_2, a_3, \dots\}$ so that $\Pr(U = a_1) \geq \Pr(U = a_2) \geq \dots$, and assign the binary sequences $\lambda, 0, 1, 00, 01, 10, 11, ...$ to the letters $a_1, a_2, \dots$ in that order. Now observe that the i'th binary sequence in the list $\lambda, 0, 1, 00, 01, \dots$ is of length $\lfloor \log_2 i \rfloor$.]