The Mathematics of Machine Learning
Final Exam

December 10, 2020

The final exam is due on December 21st. Please make sure to show all work so that you can get partial credit. Also, please list any parts of definitions, properties or theorems required for the problems. When questions ask to bound an expressions, provide the tightest bounds possible within the confines of the question. This exam will count as 40% of your grade. The total number of points is 400, representing that it is worth 4 times one of the homework assignments.

1. (100 points total) For this problem, consider the following setup:

   - Let $Z_1, Z_2, \dots, Z_m$ be IID random variables with $\mathbb{P}(Z_i = 1) = \mathbb{P}(Z_i = 0) = \frac{1}{2}$.
   - Define the randomly generated function, $f : (0, 1] \to \{0, 1\}$ such that $f(x)$ is Z_i for $x \in \left(\frac{i-1}{m}, \frac{i}{m}\right]$, that is, $f(x) = Z_{\lceil mx \rceil}$.
   - Let $X_1, X_2, \dots$ be IID, independent of the Z's and uniformly distributed on $(0, 1]$.
   - Let $Y_i = f(X_i)$.

   (a) (20 points) We remind the reader of the definition of continuity: $f : (0, 1] \to \mathbb{R}$ is **continuous** if for all $x \in (0, 1]$ and all $\epsilon > 0$, there is a $\delta > 0$ such that for all y with $|x - y| < \delta$:

   $$|f(x) - f(y)| < \epsilon$$

   What is the probability that the function f is continuous? (1/2)^m

   (b) (20 points) What is the probability that $X_1 \in \left(\frac{i-1}{m}, \frac{i}{m}\right]$? p = Delta = (1/m)

   (c) (20 points) What is the probability that $\{X_1, X_2, \dots, X_n\} \cap \left(\frac{i-1}{m}, \frac{i}{m}\right] \neq \varnothing$?

   (d) (20 points) What is the probability that $\{X_1, X_2, \dots, X_n\} \cap \left(\frac{\lceil mX_{n+1} \rceil - 1}{m}, \frac{\lceil mX_{n+1} \rceil}{m}\right] \neq \varnothing$?

   (e) (20 points) Consider an the estimator given by:

Yi和Yn+1的关系?
即使右边是选择了Yi,
和Yn+1有什么关系?

   $$\hat{Y}_{n+1} = \begin{cases} Y_i & \text{if } X_i \in \left(\frac{\lceil mX_{n+1} \rceil - 1}{m}, \frac{\lceil mX_{n+1} \rceil}{m}\right] \text{ for some } i \in \{1, 2, \dots, n\} \\ \frac{1}{2} & \text{otherwise} \end{cases}$$

   Note that $\hat{Y}_{n+1}$ can be seen to be well-defined since $Y_i = Y_j$ if $X_i, X_j \in \left(\frac{\lceil mX_{n+1} \rceil - 1}{m}, \frac{\lceil mX_{n+1} \rceil}{m}\right]$. What is the probability that $\hat{Y}_{n+1} = Y_{n+1}$?

2. (80 points total) You manage money for a company that provides you with capital and takes profits in the following way:

   - If your return is positive in a given period, your capital is increased by α times the return. For example, if you make 10% in one period and $\alpha = \frac{1}{2}$, you would then be managing 1.05 times the amount you had previously managed.
   - If your return is negative in a given period, your capital is decreased by β times the negative of the return. For example, if you lose 10% in one period and $\beta = 2$, you would then be managing 0.8 times the amount you had previously managed.

   Let w be portfolio weights, that is, the vector of percentages of your captial invested in each asset and R_t the IID vector of returns in each asset in each period. You can assume that R_t is sufficiently well-behaved in terms of the existence of any moments required by the following problems. Let C_t be the amount of capital managed after period t and Let $C_0 = 1$.

   (a) (20 points) What is the value of:

   $$\lim_n \frac{1}{n} \ln(C_n)$$

   in terms of α, β, R_1 and w? You may also use expected values of arbitrary functions of these variables including the positive part function given by $x^+ = \max(x, 0)$ and the negative part function given by $x^- = \max(-x, 0)$.

   (b) (28 points) You can invest some portion of your capital in a risky stock. In each period, there are two possible scenarios: scenario 1, which occurs with probability p_1 and scenario 2, which occurs with probability $1 - p_1$. If scenario i occurs, then $R_t = r_i$, that is, the return on the risky stock is r_i. Let w be the fraction of capital invested in the risky stock. What is the value of w which optimizes the long-term growth in capital in terms of α, β, p_1, r_1 and r_2 assuming that $r_2 < 0 < r_1$?

   (c) (32 points total; 2 points each) Let $r_1 = 1.0$ (100% return) and $r_2 = -0.5$ (50% loss) and $p_1 = 0.5$. Find the portfolio, w^*, corresponding to optimal long-term growth for each of the following α and β:

$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	$\boldsymbol{w^*}$
$\frac{1}{100}$	$\frac{1}{100}$	
$\frac{1}{2}$	1	
$\frac{1}{2}$	2	
1	1	

3. (100 points total) Define an m-layer neural network as a function $X_{m,1} : \mathbb{R}^d \to \mathbb{R}$ as follows:

- Let $X_{0,i} : \mathbb{R}^d \to \mathbb{R}$ for $i \in \{1, 2, \ldots, d\}$ be the input $X_{0,i}(x) = x_i$.
- Let $X_{j,i}$, for $j \in \{1, \ldots, j\}$ and $i \in \{1, 2, \ldots, k_j\}$, be the *ith* element of the *jth* layer defined as follows:

$$X_{j,i}(x) = f\left(a_{j,i}^T X_{j-1}(x)\right)$$

where:

$$\begin{aligned}
f(x) &= \frac{1}{1+\exp(-x)} \text{ (the logistic function)} \\
X_{j-1}(x) &= \left(X_{j-1,1}(x), X_{j-1,2}(x), \ldots, X_{j-1,k_{j-1}}(x)\right) \\
a_{j,i} &= \left(a_{j,i,1}, a_{j,i,2}, \ldots, a_{j,i,k_j}\right)
\end{aligned}$$

Note that k_j is the number of elements in the *jth* layer.

(a) (24 points) What is the Lipschitz constant of the logistic function $f(x)$?

(b) (28 points) For arbitrary differentiable functions $g_1 : \mathbb{R} \to \mathbb{R}$ and $g_2 : \mathbb{R} \to \mathbb{R}$, with Lipschitz constants C_1 and C_2, respectively, provide a bound on the Lipschitz constant of the function h defined by $h(x) = g_1(g_2(x))$.

(c) (24 points) Let the hidden layer elements $X_{j,i}(x)$ for a fixed $j \in \{1, 2, \ldots, m-1\}$ have Lipschitz constants $C_{j,i}$ and let $C_j = \left(C_{j,1}, C_{j,2}, \ldots, C_{j,k_j}\right)$. The L_∞ norm of a vector u is defined as:

$$\|u\|_\infty = \max_i |u_i|$$

What is a bound on the Lipschitz constant of $X_{j+1,i}$ in terms of $\|a_{j+1,i}\|_\infty$ and $\|C_j\|_\infty$? You can use the following fact:

$$\|u^T v\|_\infty \leq \|u\|_\infty \|v\|_\infty$$

for any vectors u and v.

(d) (24 points) If $\|a_{j,i}\| \leq a$ for a constant a, what is a bound for the Lipschitz constant of $X_{m,1}(x)$ in terms of a and m?

4. (120 points total) Let F_C be the set of all Lipschitz functions $f : [0,1]^d \to [0,1]$. We define the L_∞ norm on F_C as:

$$\|f\|_\infty = \sup_{x \in [0,1]^d} |f(x)|$$

It can be shown that the covering number of F_C using the L_∞ norm is bounded by:

$$\mathcal{N}(\epsilon, F_C, \|\cdot\|_\infty) \leq \left(\frac{8}{\epsilon}\right)^{\left(\left(\frac{2C}{\epsilon}\right)^d\right)}$$

Let $(X_1, Y_1), (X_2, Y_2), \ldots$ be IID random vectors with $X_i \in [0,1]^d$.

(a) (40 points) Use Hoeffding's inequality to come up with a bound on:

$$\mathbb{P}\left(\sup_{f \in F_C} \hat{E_n}[f(X)] - E[f(X)] \geq \epsilon\right)$$

for any $\epsilon > 0$.

(b) (40 points) Now let $L : \mathbb{R} \times \mathbb{R} \to [0,1]$ be a loss function which is in F_C. Use the bound from the previous part of this problem to bound:

$$\mathbb{P}\Big(E\Big[L\Big(Y, \hat{f}(X)\Big)\Big] - E[L(Y, f^*(X))] \geq \epsilon\Big) \tag{1}$$

for any $\epsilon > 0$ where f^* is the optimal estimator:

$$f^* = \operatorname*{argmin}_f E[L(Y, f(X))]$$

and $\hat{f}$ is the optimal estimator based on the observed data:

$$\hat{f} = \operatorname*{argmin}_f \hat{E_n}[L(Y, f(X))]$$

(c) (40 points) Use the previous bound to determine the number of samples needed so that the probability in Formula (1) is at most δ.