18.657. Fall 2105
Rigollet September 27, 2015

Problem set #1 (due Wed., October 7)

Problem 1. Discriminant analysis

Let $(X, Y) \in \mathbb{R}^d \times \{0, 1\}$ be a random pair such that $\mathbb{P}(Y = k) = \pi_k > 0$ ($\pi_0 + \pi_1 = 1$) and the conditional distribution of X given Y is $X|Y \sim \mathcal{N}(\mu_Y, \Sigma_Y)$, where $\mu_0 \neq \mu_1 \in \mathbb{R}^d$ and $\Sigma_0, \Sigma_1 \in \mathbb{R}^{d\times d}$ are mean vectors and covariance matrices respectively.

1. What is the (unconditional) density of X?
2. Assume that $\Sigma_0 = \Sigma_1 = \Sigma$ is a positive definite matrix. Compute the Bayes classifier h^* as a function of $\mu_0, \mu_1, \pi_0, \pi_1$ and Σ. What is the nature of the sets $\{h^* = 0\}$ and $\{h^* = 1\}$?
3. Assume now that $\Sigma_0 \neq \Sigma_1$ are two positive definite matrices. What is the nature of the sets $\{h^* = 0\}$ and $\{h^* = 1\}$?

Problem 2. VC dimensions

1. Let $\mathcal{C}$ be the class of convex polygons in $\mathbb{R}^2$ with d vertices. Show that $\mathsf{VC}(\mathcal{C}) = 2d + 1$.
2. Let $\mathcal{C}$ be the class of convex compact sets in $\mathbb{R}^2$. Show that $\mathsf{VC}(\mathcal{C}) = \infty$.
3. Let $\mathcal{C}$ be finite. Show that $\mathsf{VC}(\mathcal{C}) \leq \log_2(\text{card}\, \mathcal{C})$.
4. Give an example of a class $\mathcal{C}$ such that $\text{card}\, \mathcal{C} = \infty$ and $\mathsf{VC}(\mathcal{C}) = 1$.

Problem 3. Glivenko-Cantelli Theorem

Let $X_1, \ldots, X_n$ be n i.i.d copies of X that has cumulative distribution function (cdf) $F(t) = \mathbb{P}(X \leq t)$. The *empirical* cdf of X is defined by

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq t)\,.$$

1. Compute the mean and the variance of $\hat{F}_n(t)$ and conclude that $\hat{F}_n(t) \to F(t)$ as $n \to \infty$ almost surely (hint: use Borel-Cantelli).

2. Show that for $n \geq 2$

$$\sup_{t \in \mathbb{R}} \left|\hat{F}_n(t) - F(t)\right| \leq C\sqrt{\frac{\log(n/\delta)}{n}}$$

with probability $1 - \delta$.

Problem 4. Concentration

1. Let $X_1, \ldots, X_n$ be n i.i.d copies of $X \in [0,1]$. Each X_i represents the size of a packages to be shipped. The shipping containers are bins of size 1 (so that each bin can hold a set of packages whose sizes sum to at most 1). Let B_n be the minimal number of bins needed to store the n packages. Show that

$$\mathbb{P}(|B_n - \mathbb{E}[B_n]| \geq t) \leq 2e^{-\frac{2t^2}{n}}\,.$$

2. Let $X_1, \ldots, X_n$ be n i.i.d copies of $X \in \mathbb{R}^d$, $\mathbb{E}[X] = 0$ and assume that $\|X_i\| \leq 1$ almost surely for all i. Let $\bar{X}$ denote the average of the X_is. Prove the following inequalities (the constant C may change from one inequality to the other)

$$(a)\ \mathbb{P}\Big[\big\|\bar{X}\big\| - \mathbb{E}\big\|\bar{X}\big\| \geq t\Big] \leq e^{-Cnt^2}\,, \quad (b)\ \mathbb{E}\big\|\bar{X}\big\| \leq \frac{C}{\sqrt{n}}\,, \quad (c)\ \mathbb{P}\Big[\big\|\bar{X}\big\| \geq t\Big] \leq 2e^{-Cnt^2}$$

3. Let $X_1, \ldots, X_n$ be n iid random variables, i.e. such that X_i and $-X_i$ have the same distribution. Let $\bar{X}$ denote the average of the X_is and $V = n^{-1}\sum_{i=1}^n X_i^2$. Show that

$$\mathbb{P}\Big[\frac{\bar{X}}{\sqrt{V}} > t\Big] \leq e^{-\frac{nt^2}{2}}\,.$$

[Hint: introduce Rademacher random variables].

MIT OpenCourseWare
http://ocw.mit.edu

18.657 Mathematics of Machine Learning
Fall 2015

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.