

Homework #0

CSE 546: Machine Learning

Michael Ross

Due: 10/4/18 11:59 PM

1 Analysis

1. [1 points] A set $A \subseteq \mathbb{R}^n$ is *convex* if $\lambda x + (1 - \lambda)y \in A$ for all $x, y \in A$ and $\lambda \in [0, 1]$. A *norm* $\|\cdot\|$ over \mathbb{R}^n is defined by the properties: i) non-negative: $\|x\| \geq 0$ for all $x \in \mathbb{R}^n$ with equality if and only if $x = 0$, ii) absolute scalability: $\|ax\| = |a| \|x\|$ for all $a \in \mathbb{R}$ and $x \in \mathbb{R}^n$, iii) triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

- a. Using just the definitions above, show that the set $\{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is convex for any norm $\|\cdot\|$.

Answer:

Let $A = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ and $x, y \in A$

$$\begin{aligned}\|\lambda x + (1 - \lambda)y\| &= \|\lambda x\| + \|(1 - \lambda)y\| = |\lambda| \|x\| + |1 - \lambda| \|y\| \\ |\lambda| &\in [0, 1] \text{ by norm definition} \\ \|x\| &\in [0, 1] \text{ by norm and set definitions} \\ \|y\| &\in [0, 1] \text{ by norm and set definitions} \\ \implies \|\lambda x + (1 - \lambda)y\| &\leq 1\end{aligned}$$

$$\begin{aligned}\lambda x + (1 - \lambda)y &\text{ is a linear combination of elements of } \mathbb{R}^n \\ \implies \lambda x + (1 - \lambda)y &\in \mathbb{R}^n\end{aligned}$$

Thus $\lambda x + (1 - \lambda)y \in A$ for any x, y

- b. Show that $(\sum_{i=1}^n |x_i|^{1/2})^2$ is or is not a norm.

Answer:

$$\begin{aligned}\text{i) } x_i &\in \mathbb{R} \\ \implies |x_i|^{1/2} &\in \mathbb{R} \\ \implies \sum_{i=1}^n |x_i|^{1/2} &\in \mathbb{R} \\ \implies (\sum_{i=1}^n |x_i|^{1/2})^2 &\geq 0\end{aligned}$$

$$\begin{aligned}\text{ii) } \|ax\| &= (\sum_{i=1}^n |ax_i|^{1/2})^2 \\ &= (\sum_{i=1}^n (|a||x_i|)^{1/2})^2 \\ &= (\sum_{i=1}^n |a|^{1/2}|x_i|^{1/2})^2 \\ &= (|a|^{1/2} \sum_{i=1}^n |x_i|^{1/2})^2 \\ &= |a| (\sum_{i=1}^n |x_i|^{1/2})^2 \\ &= |a| \|x\|\end{aligned}$$

iii)

2. [1 points] For any $x \in \mathbb{R}^n$, define the following norms: $\|x\|_1 = \sum_{i=1}^n |x_i|$, $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$, $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$. Show that $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$.

3. [1 points] For possibly non-symmetric $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}$, let $f(x, y) = x^T \mathbf{A}x + y^T \mathbf{B}y + c$. Define $\nabla_z f(x, y) = \begin{bmatrix} \frac{\partial f(x, y)}{\partial z_1} & \frac{\partial f(x, y)}{\partial z_2} & \dots & \frac{\partial f(x, y)}{\partial z_n} \end{bmatrix}^T$. What is $\nabla_x f(x, y)$ and $\nabla_y f(x, y)$?

4. [1 points] Let \mathbf{A} and \mathbf{B} be two $\mathbb{R}^{n \times n}$ symmetric matrices. Suppose \mathbf{A} and \mathbf{B} have the exact same set of eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ with the corresponding eigenvalues $\alpha_1, \alpha_2, \dots, \alpha_n$ for \mathbf{A} , and $\beta_1, \beta_2, \dots, \beta_n$ for \mathbf{B} . Please write down the eigenvectors and their corresponding eigenvalues for the following matrices:

a. $\mathbf{C} = \mathbf{A} + \mathbf{B}$

Answer:

Eigenvectors: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ Eigenvalues: $\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_n + \beta_n$

b. $\mathbf{D} = \mathbf{A} - \mathbf{B}$

Answer:

Eigenvectors: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ Eigenvalues: $\alpha_1 - \beta_1, \alpha_2 - \beta_2, \dots, \alpha_n - \beta_n$

c. $\mathbf{E} = \mathbf{A}\mathbf{B}$

Answer:

Eigenvectors: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ Eigenvalues: $\alpha_1\beta_1, \alpha_2\beta_2, \dots, \alpha_n\beta_n$

d. $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$ (assume \mathbf{A} is invertible)

Answer:

Eigenvectors: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ Eigenvalues: $\beta_1/\alpha_1, \beta_2/\alpha_2, \dots, \beta_n/\alpha_n$

5. [1 points] A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is *positive-semidefinite (PSD)* if $x^T \mathbf{A} x \geq 0$ for all $x \in \mathbb{R}^n$.

a. For any $y \in \mathbb{R}^n$, show that yy^T is PSD.

Answer:

$$x^T yy^T x = (x \cdot y)(y \cdot x) = (x \cdot y)^2 \geq 0$$

b. Let X be a random vector in \mathbb{R}^n with covariance matrix $\Sigma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$. Show that Σ is PSD.

Answer:

$$\begin{aligned} x^T \Sigma x &= x^T \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T] x \\ &= \mathbb{E}[x^T (X - \mathbb{E}[X])(X - \mathbb{E}[X])^T x] \text{ since } x \text{ is not random } x\mathbb{E}[X] = \mathbb{E}[xX] \\ \text{Let } \gamma &= (X - \mathbb{E}[X])^T x = x^T (X - \mathbb{E}[X]) \\ x^T \Sigma x &= \mathbb{E}[\gamma^2] \\ \gamma \in \mathbb{R} &\implies \gamma^2 \geq 0 \implies \mathbb{E}[\gamma^2] \geq 0 \\ &\implies x^T \Sigma x \geq 0 \end{aligned}$$

c. Assume \mathbf{A} is a symmetric matrix so that $\mathbf{A} = \mathbf{U} \text{diag}(\alpha) \mathbf{U}^T$ where $\text{diag}(\alpha)$ is an all zeros matrix with the entries of α on the diagonal and $\mathbf{U}^T \mathbf{U} = \mathbf{I}$. Show that \mathbf{A} is PSD if and only if $\min_i \alpha_i \geq 0$. (Hint: compute $x^T \mathbf{A} x$ and consider values of x proportional to the columns of \mathbf{U} , i.e., the orthonormal eigenvectors).

Answer:

$$\begin{aligned} x^T \mathbf{A} x &= x^T \mathbf{U} \text{diag}(\alpha) \mathbf{U}^T x \\ \text{Let } x &= \mathbf{U} \beta \text{ where } \beta \in \mathbb{R}^n \\ x^T \mathbf{A} x &= \beta^T \mathbf{U}^T \mathbf{U} \text{diag}(\alpha) \mathbf{U}^T \mathbf{U} \beta \\ &= \beta^T \text{diag}(\alpha) \beta \\ \text{Switching to index notation} \\ &= \beta_i \text{diag}(\alpha)_{ij} \beta_j \\ \text{diag}(\alpha)_{ij} &= 0 \text{ for } i \neq j \\ &\implies x^T \mathbf{A} x = \sum_{i=1}^n \beta_i \beta_i \alpha_i \end{aligned}$$

6. [1 points] Let X and Y be real independent random variables with PDFs given by f and g , respectively. Let h be the PDF of the random variable $Z = X + Y$.

a. Derive a general expression for h in terms of f and g

b. If X and Y are both independent and uniformly distributed on $[0, 1]$ (i.e. $f(x) = g(x) = 1$ for $x \in [0, 1]$ and 0 otherwise) what is h , the PDF of $Z = X + Y$?

c. For these given explicit distributions, what is $\mathbb{P}(X \leq 1/2 | X + Y \geq 5/4)$?

7. [1 points] A random variable $X \sim \mathcal{N}(\mu, \sigma^2)$ is Gaussian distributed with mean μ and variance σ^2 . Given that for any $a, b \in \mathbb{R}$, we have that $Y = aX + b$ is also Gaussian, find a, b such that $Y \sim \mathcal{N}(0, 1)$.

Answer:

$$\mathbb{E}[Y] = 0$$

$$a\mathbb{E}[X] + b = 0$$

$$a\mu + b = 0$$

$$\mu = -b/a$$

$$\mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 1$$

$$\text{Since } \mathbb{E}[Y] = 0$$

$$\mathbb{E}[Y^2] = 1$$

$$\mathbb{E}[a^2 X^2 + 2abX + b^2] = 1$$

$$a^2 \mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 = 1$$

$$a^2 \mathbb{E}[X^2] - a^2 \mathbb{E}[X]^2 + a^2 \mathbb{E}[X]^2 + 2ab\mathbb{E}[X] + b^2 = 1$$

$$a^2 \sigma^2 + a^2 \mu^2 + 2ab\mu + b^2 = 1$$

$$\text{Plug } \mu = -b/a \text{ in}$$

$$a^2 \sigma^2 + b^2 - 2b^2 + b^2 = 1$$

$$\implies a = 1/\sigma$$

$$\implies b = -\mu/\sigma$$

8. [1 points] If $f(x)$ is a PDF, we define the cumulative distribution function (CDF) as $F(x) = \int_{-\infty}^x f(y)dy$. For any function $g : \mathbb{R} \rightarrow \mathbb{R}$ and random variable X with PDF $f(x)$, define the expected value of $g(X)$ as $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(y)f(y)dy$. For a boolean event A , define $\mathbf{1}\{A\}$ as 1 if A is true, and 0 otherwise. Thus, $\mathbf{1}\{x \leq a\}$ is 1 whenever $x \leq a$ and 0 whenever $x > a$. Note that $F(x) = \mathbb{E}[\mathbf{1}\{X \leq x\}]$. Let X_1, \dots, X_n be independent and identically distributed random variables with CDF $F(x)$. Define $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}$.

a. For any x , what is $\mathbb{E}[\hat{F}_n(x)]$?

Answer:

$$\begin{aligned} \mathbb{E}[\hat{F}_n(x)] &= \mathbb{E}[\frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\mathbf{1}\{X_i \leq x\}] \\ &= \frac{1}{n} \sum_{i=1}^n F(x) \\ &= F(x) \end{aligned}$$

b. For any x , show that $\mathbb{E}[(\hat{F}_n(x) - F(x))^2] = \frac{F(x)(1-F(x))}{n}$

Answer:

$$\begin{aligned} \mathbb{E}[(\hat{F}_n(x) - F(x))^2] &= \mathbb{E}[\hat{F}_n(x)^2 - 2\hat{F}_n(x)F(x) + F(x)^2] \\ &= \mathbb{E}[\hat{F}_n(x)^2] - 2\mathbb{E}[\hat{F}_n(x)]F(x) + F(x)^2 \end{aligned}$$

$$\begin{aligned} \hat{F}_n(x)^2 &= \frac{1}{n^2} (\sum_{i=1}^n \mathbf{1}\{X_i \leq x\})^2 \\ &= \frac{1}{n^2} (\sum_{i=1}^n \mathbf{1}\{X_i \leq x\} \mathbf{1}\{X_1 \leq x\} + \dots + \sum_{i=1}^n \mathbf{1}\{X_i \leq x\} \mathbf{1}\{X_n \leq x\}) \\ &= \frac{1}{n^2} (\sum_{i=1}^n \mathbf{1}\{X_i \leq x\} + \dots + \sum_{i=1}^n \mathbf{1}\{X_i \leq x\}) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{X_i \leq x\} \\ &= \hat{F}_n(x) \end{aligned}$$

$$\begin{aligned} \mathbb{E}[(\hat{F}_n(x) - F(x))^2] &= F(x) - 2F(x)F(x) + F(x)^2 \\ &= F(x)(1 - F(x)) \end{aligned}$$

c. Using part b., show that $\sup_{x \in \mathbb{R}} \mathbb{E}[(\hat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$.

Answer:

$$\sup_{x \in \mathbb{R}} \mathbb{E}[(\hat{F}_n(x) - F(x))^2] = \sup_{x \in \mathbb{R}} \frac{1}{n} (F(x) - F(x)^2)$$

$$\begin{aligned} \partial_{F(x)} (\frac{1}{n} (F(x) - F(x)^2)) &= \frac{1}{n} (1 - 2F(x)) \\ \arg \max_{F(x)} (\frac{1}{n} (F(x) - F(x)^2)) &= 1/2 \end{aligned}$$

$$\max \left(\frac{1}{n} (F(x) - F(x)^2) \right) = \frac{1}{4n}$$

$$\implies \sup_{x \in \mathbb{R}} \mathbb{E}[(\hat{F}_n(x) - F(x))^2] \leq \frac{1}{4n}$$

2 Programming

9. [2 points] Two random variables X and Y have equal distributions if their CDFs, F_X and F_Y , respectively, are equal: $\sup_x |F_X(x) - F_Y(x)| = 0$. The central limit theorem says that the sum of k independent, zero-mean, variance-1/ k random variables converges to a Gaussian distribution as k goes off to infinity. We will study this phenomenon empirically (you will use the Python packages Numpy and Matplotlib). Define $Y^{(k)} = \frac{1}{\sqrt{k}} \sum_{i=1}^k B_i$ where each B_i is equal to -1 and 1 with equal probability. It is easy to verify (you should) that $\frac{1}{\sqrt{k}} B_i$ is zero-mean and has variance $1/k$.

- a. For $i = 1, \dots, n$ let $Z_i \sim \mathcal{N}(0, 1)$. If $F(x)$ is the true CDF from which each Z_i is drawn (i.e., Gaussian) and $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{Z_i \leq x\}$, use the homework problem above to choose n large enough such that $\sup_x \sqrt{\mathbb{E}[(\hat{F}_n(x) - F(x))^2]} \leq 0.0025$, and plot $\hat{F}_n(x)$ from -3 to 3 . (Hint: use `Z=np.random.randn(n)` to generate the random variables, and `import matplotlib.pyplot as plt`; `plt.step(sorted(Z), np.arange(1,n+1)/float(n))` to plot).

Answer:

```
import matplotlib.pyplot as plt
import numpy as np

supMax=0.0025
n=int(1.0/(2.0*supMax)**2)
Z=np.random.randn(n)

plt.step(sorted(Z),np.arange(1,n+1)/float(n))
plt.xlabel("Observations")
plt.ylabel("Probability")
plt.xlim((-3,3))
plt.ylim((0,1))
plt.grid()
plt.show()
```

- b. For each $k \in \{1, 8, 64, 512\}$ generate n independent copies $Y^{(k)}$ and plot their empirical CDF on the same plot as part a. (Hint: you can use `np.sum(np.sign(np.random.randn(n, k))*np.sqrt(1./k), axis=1)` to generate n of the $Y^{(k)}$ random variables.)

Answer:

```
import matplotlib.pyplot as plt
import numpy as np

supMax=0.0025
n=int(1.0/(2.0*supMax)**2)
Z=np.random.randn(n)

k=[1,8,64,512]
i=1000
for j in range(len(k)):
    x=np.sum(np.sign(np.random.randn(i, k[j]))*np.sqrt(1./k[j]), axis=1)
    plt.step(sorted(x),np.arange(1,i+1)/float(i))

plt.step(sorted(Z),np.arange(1,n+1)/float(n))
```

```
plt.xlabel("Observations")
plt.ylabel("Probability")
plt.xlim((-3,3))
plt.ylim((0,1))
k.append('Gaussian')
plt.legend(k)
plt.grid()
plt.show()
```

Be sure to always label your axes. Your plot should look something like the following (Tip: checkout **seaborn** for instantly better looking plots.)

