EECS 545 – Machine Learning - Homework #1

Due: 17:00 9/22/2017

Homework Policy: Working in groups is allowed, but each member must submit their own writeup. Please write the members of your group on your solutions (maximum allowed group size is 4). Homeworks will be submitted via Gradescope (http://gradescope.com/) as a pdf file (including your code).

1) **Linear Algebra (20 pts).**

(a) Are the following statements **true** or **false**? If true, prove it; if false, show a counterexample.

(i) The inverse of a symmetric matrix is itself symmetric.

(ii) All 2×2 orthogonal matrices have the following form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ or } \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}.$$

(iii) Let $A = \begin{bmatrix} -8 & -1 & -6 \\ -3 & -5 & -7 \\ -4 & -9 & -2 \end{bmatrix}$. A can be written as $A = CC^T$ for some matrix C.

2) **Probability (20 pts).**

(a) Random variables X and Y have a joint distribution $p(x, y)$. Prove the following results. You can assume continuous distributions for simplicity.

(i) $\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$

(ii) $\mathbb{E}[I[X \in \mathcal{C}]] = P(X \in \mathcal{C})$, where $I[X \in \mathcal{C}]$ is the indicator function[1] of an arbitrary set $\mathcal{C}$.

(iii) $\text{var[X]} = \mathbb{E}_Y[\text{var}_X[X|Y]] + \text{var}_Y[\mathbb{E}_X[X|Y]]$

(iv) If X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

(v) If X and Y take values in $\{0, 1\}$ and $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$, then X and Y are independent.

(b) For the following equations, describe the relationship between them. Write one of four answers: "=", "$\leq$", "$\geq$", or "depends" to replace the "?". Choose the most specific relation that always holds and briefly explain why. Assume all probabilities are non-zero.

[1] $I[X \in \mathcal{C}] := 1$ if $X \in \mathcal{C}$ and 0 otherwise

(i) $P(H = h, D = d)$? $P(H = h)$

(ii) $P(H = h|D = d)$? $P(H = h)$

(iii) $P(H = h|D = d)$? $P(D = d|H = h)P(H = h)$

3) **Positive (Semi-)Definite Matrices (20 pts).** Let A be a real, *symmetric* $d \times d$ matrix. We say A is *positive semi-definite* (PSD) if, for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^\top A\mathbf{x} \geq 0$. We say A is *positive definite* (PD) if, for all $\mathbf{x} \neq 0$, $\mathbf{x}^\top A\mathbf{x} > 0$. We write $A \succeq 0$ when A is PSD, and $A \succ 0$ when A is PD.

The *spectral theorem* says that every real symmetric matrix A can be expressed $A = U\Lambda U^\top$, where U is a $d \times d$ matrix such that $UU^\top = U^\top U = I$ (called an orthogonal matrix), and $\Lambda = diag(\lambda_1, ..., \lambda_d)$. Multiplying on the right by U we see that $AU = U\Lambda$. If we let $\mathbf{u}_i$ denote the i^{th} column of U, we have $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for each i. This expression reveals that the λ_i are eigenvalues of A, and the corresponding columns $\mathbf{u}_i$ are eigenvectors associated to λ_i.

Using the spectral decomposition, show that

(a) A is PSD iff $\lambda_i \geq 0$ for each i.

(b) A is PD iff $\lambda_i > 0$ for each i.

Hint: Use the following representation

$$U\Lambda U^T = \sum_{i=1}^{d} \lambda_i \mathbf{u_i}\mathbf{u_i}^T .$$

4) **Optimization (20 pts).** Recall that a function f is *convex* if

$$f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y})$$

for all $t \in [0, 1]$ and $\mathbf{x}, \mathbf{y}$ in the domain of f. A function f is called *strictly convex* if the above relation is a strict inequality for all $\mathbf{x} \neq \mathbf{y}$ in the domain of f. A function $f(\mathbf{x})$ is concave iff $-f(\mathbf{x})$ is convex.

(a) Show that the affine function $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ is convex and concave. Is $f(\mathbf{x})$ strictly convex ?

(b) Show that if f is strictly convex, then f has at most one global minimizer.

For the next two parts, the following fact will be helpul. A twice continuously differentiable function admits the quadratic expansion

$$f(\mathbf{x}) = f(\mathbf{y}) + \langle \triangledown f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{1}{2}\langle \mathbf{x} - \mathbf{y}, \triangledown^2 f(\mathbf{y})(\mathbf{x} - \mathbf{y}) \rangle + o(\|\mathbf{x} - \mathbf{y}\|^2) \quad (1)$$

where $o(t)$ denotes a function satisfying $\lim_{t\to 0} \frac{o(t)}{t} = 0$, as well as the expansion

$$f(\mathbf{x}) = f(\mathbf{y}) + \langle \triangledown f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{1}{2}\langle \mathbf{x} - \mathbf{y}, \triangledown^2 f(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))(\mathbf{x} - \mathbf{y}) \rangle \quad (2)$$

for some $t \in (0,1)$.

(c) Show that if f is twice continuously differentiable and $\mathbf{x}^*$ is a local minimizer, then $\bigtriangledown^2 f(\mathbf{x}^*) \succeq 0$, *i.e.*, the Hessian of f is positive semi-definite at the local minimizer $\mathbf{x}^*$.

(d) Show that if f is twice continuously differentiable, then f is convex if and only if the Hessian $\bigtriangledown^2 f(\mathbf{x})$ is positive semi-definite for all $\mathbf{x} \in \mathbb{R}^d$.

(e) Consider the function $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \mathbf{b}^\top\mathbf{x} + c$, where A is a symmetric $d \times d$ matrix. Derive the Hessian of f. Under what conditions on A is f convex? Strictly convex?

5) **Programming (20 pts).**

Singular value decomposition (SVD) factorizes a $m \times n$ matrix X as $X = U\Sigma V^\top$, where $U \in \mathbb{R}^{m\times m}$ and $U^\top U = UU^\top = I$, $\Sigma \in \mathbb{R}^{m\times n}$ contains non-increasing non-negative values along its diagonal and zeros elsewhere, and $V \in \mathbb{R}^{n\times n}$ and $V^\top V = VV^\top = I$. Hence matrix can be represented as $X = \sum_{i=1}^{d} \sigma_i u_i v_i^T$ where u_i denotes the i^{th} column of U and v_i denotes the i^{th} column of V. Download the **image**[2] and load it as the matrix X (in greyscale).

(a) Perform SVD on this matrix and zero out all but top k singular values to form an approximation $\tilde{X}$. Specifically, compute $\tilde{X} = \sum_{i=1}^{k} \sigma_i u_i v_i^T$, display the resulting approximation $\tilde{X}$ as an image, and report $\frac{\|X-\tilde{X}\|_F}{\|X\|_F}$ for $k \in \{2, 10, 40\}$.

(b) How many numbers do you need to describe the approximation $X = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ for $k \in \{2, 10, 40\}$?

Hint:
Matlab can load the greyscale image via
```
X=double(rgb2gray(imread('harvey-saturday-goes7am.jpg')))
```
and display the image using `imagesc()`

In Python 2.7 you can use:
```
import numpy, Image
X=numpy.asarray(Image.open('harvey.jpg').convert('L'))
```

[2]https://www.nasa.gov/sites/default/files/thumbnails/image/harvey-saturday-goes7am.jpg
(NOAA's GOES-East satellite's image of Hurricane Harvey in the western Gulf of Mexico on Aug. 26, 2017. Credit: NASA/NOAA GOES Project)