Homework 2
MATH 123 - Fall 2018
Tufts University, Department of Mathematics
Due: September 18, 2018

1. Question 1

Let $\Sigma \in \mathbb{R}^{d\times d}$ be symmetric. Let $F : \mathbb{R}^d \to \mathbb{R}$ be the function $F(u) = u^T \Sigma u$, where u is understood as a column vector, i.e. as $u \in \mathbb{R}^{d\times 1}$. Show

$$\frac{\partial F}{\partial u} = 2\Sigma u.$$

2. Question 2

Recall that the *variance* of a set of numbers $x_1, \dots x_n \in \mathbb{R}$ is $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$, where $\mu = \frac{1}{n}\sum_{i=1}^n x_i$ is the mean. For each of the following statements, prove or give a counterexample.

(a) The variance is *translation invariant*, i.e. the variance of $x_1, \dots, x_n$ is the same as the variance of the translated set $x_1 + T, \dots, x_n + T$ for any fixed $T \in \mathbb{R}$.
(b) The variance is 0 if and only if $x_i = C$, $\forall i = 1, \dots, n$ for some constant C. In other words, the variance is 0 if and only if all data points are equal.
(c) The variance is *additive*, i.e. if $x_1, \dots, x_n$ have variance σ_x^2 and $y_1, \dots y_m$ have variance σ_y^2, then the concatenated set $x_1, \dots, x_n, y_1, \dots, y_m$ has variance $\sigma_x^2 + \sigma_y^2$.

3. Question 3

A matrix $A \in \mathbb{R}^d$ is said to be *positive semi-definite* if $y^T Ay \geq 0$ for all $y \in \mathbb{R}^{d\times 1}$. The matrix A is said to be *positive definite* if it is positive semi-definite and $y^T Ay = 0$ if and only if $y = (0, 0, \dots, 0)$.

(a) Let $x_1, \dots, x_n \in \mathbb{R}^{1\times d}$ be data. Let $\Sigma = \frac{1}{n}\sum_{i=1}^n x_i^T x_i$ be the covariance matrix. Prove Σ is positive semi-definite.
(b) Is Σ necessarily positive definite?

4. Question 4

When dimension reducing data in $\mathbb{R}^D$ with PCA, the choice of embedding dimension is crucial. Many heuristics exist to estimate a good dimension. One is to choose the embedding dimension d^* to be the smallest dimension such that some proportion (say, .95) of the variance of the data is preserved by projecting onto the first d principal components:

$$d^* = \min_{d'} \left\{ \frac{\sum_{i=1}^{d'} \lambda_i}{\sum_{i=1}^{d} \lambda_i} > .95 \right\}.$$

(a) Intuitively, when will this give d^* small?
(b) Intuitively, when will this give d^* big?
(c) Are there any situations in which d^* is roughly $.95 * D$?

5. Question 5

Download the corrected ‘SalinasA’ data from `http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes`.

(a) Compute the principal component decomposition of the data.
(b) How many dimensions are needed to preserve 95% of the variance in the data?
(c) Compute and display the first 3 and last 3 principal components. Are there any obvious contrasts?