CS 8850: Advanced Machine Learning **Fall 2017**

Homework 1: Review of Linear Algebra

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN DUE 8/30/2017

Problem 1.1. Show that $\mathbb{R}^{\mathrm{D}}$ is a vector space.

Problem 1.2. Vector spaces are, by definition, *closed* under linear combinations. For example, when you add or multiply elements of $\mathbb{R}^{\mathrm{D}}$, you end up with an element of $\mathbb{R}^{\mathrm{D}}$. In other words, you cannot *fall* out of $\mathbb{R}^{\mathrm{D}}$ by adding or multiplying. As we showed before, $\mathbb{R}^{\mathrm{D}}$ is a vector space, and hence it is closed under linear combinations. Vector spaces are not necessarily closed under *all* mathematical operations.

(a) Show that $\mathbb{R}^{\mathrm{D}}$ is *not* closed under square roots.

(b) Give an example of a vector space that is closed under square roots (besides being closed under linear combinations, as *all* vector spaces must be).

Problem 1.3. Let $\mathbf{u}_1, \ldots, \mathbf{u}_{\mathrm{R}} \in \mathbb{R}^{\mathrm{D}}$. Show that $\mathrm{span}[\mathbf{u}_1, \ldots, \mathbf{u}_{\mathrm{R}}]$ is a subspace.

Problem 1.4. In this problem we will generate and visualize data lying in a subspace using Matlab.

(a) Create a matrix $\mathbf{U} \in \mathbb{R}^{\mathrm{D}\times\mathrm{R}}$ whose column vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{\mathrm{R}}$ will work as a basis of our subspace.

(b) Create a matrix $\mathbf{C} \in \mathbb{R}^{\mathrm{R}\times\mathrm{N}}$ whose column vectors will work as coefficients of our data.

(c) Let $\mathbf{X} = \mathbf{UC}$. Show that the columns in $\mathbf{X}$ lie in $\mathrm{span}[\mathbf{U}]$. Hint: show that the i^{th} column in $\mathbf{X}$ is a linear combination of the columns in $\mathbf{U}$.

(d) Plot $\mathrm{N} = 100$ points in $\mathbb{R}^2$ lying in an $\mathrm{R} = 1$-dimensional subspace.

(e) Plot $\mathrm{N} = 100$ points in $\mathbb{R}^3$ lying in an $\mathrm{R} = 1$-dimensional subspace.

(f) Plot $\mathrm{N} = 1000$ points in $\mathbb{R}^3$ lying in an $\mathrm{R} = 2$-dimensional subspace.

Problem 1.5. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X} = \mathbb{R}^{\mathrm{D}}$. Show that $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^{\mathsf{T}}\mathbf{y} = \sum_{\mathrm{d}=1}^{\mathrm{D}} \mathrm{x_d y_d}$ defines an inner-product.

Problem 1.6 (Subspace fitting)**.** The file `data_2D.mat` contains a data matrix $\mathbf{X}$ whose columns $\mathbf{x}_1, \ldots, \mathbf{x}_{\mathrm{N}}$ represent points in $\mathbb{R}^2$.

(a) Find the 1-dimensional subspace (line) that best explains the data. Hint: use the *singular value decomposition* (SVD).

(b) Derive the projection operator onto this subspace.

(c) Project $\mathbf{X}$ onto this subspace.

(d) Plot the original and the projected data $\hat{\mathbf{X}}$.

(e) Compute the *mean squared error* (MSE):

$$\frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathbf{x}_{\mathrm{i}} - \hat{\mathbf{x}}_{\mathrm{i}})^2.$$

(f) Can you find an other line that produces a smaller MSE?

Similarly, the file `data_3D.mat` contains a data matrix $\mathbf{X}$ whose columns represent points in $\mathbb{R}^3$.

(g) Repeat points (a)-(e) for this dataset.

(h) Repeat points (a)-(e), but this time with a 2-dimensional subspace.

(i) Compare and discuss advantages and disadvantages of using a 1-dimensional subspace vs. a 2-dimensional subspace.

(j) What would happen if you used a 3-dimensional subspace?

(k) Which subspace would you choose?

(l) Given an arbitrary dataset, can you device a good mathematical criteria to determine the dimension of the subspace that we should use? Hint: think of the SVD.

Problem 1.7 (Iris flowers dataset)**.** The file `iris_data.mat` has a a data matrix $\mathbf{X} \in \mathbb{R}^{4\times 150}$ containing the Iris Flowers dataset (also known as Fisher's or Anderson's). The columns of $\mathbf{X}$ contain the width and length (in centimeters) of the sepals and petals of 150 samples (flowers).

(a) If you want to find a subspace that explains this data, what dimension R should this subspace have? Hint: think of the SVD.

(b) Find the R-dimensional subspace that best explains the data.

(c) Derive the projection operator onto this subspace.

(d) Project $\mathbf{X}$ onto this subspace.

(e) Compute the MSE of the projected data $\hat{\mathbf{X}}$.

(f) Can you find an other R-dimensional subspace that produces a smaller MSE?