Homework 10

Due Tuesday, December 8

Please either give the assignment to Loraine at the CDS or send it via email to the graders **before noon**.

1. *Statements (10 points).* Prove the following statements.

   a. For any subspace $\mathcal{S}$ belonging to a vector space of dimension n

   $$\dim(\mathcal{S}) + \dim\left(\mathcal{S}^{\perp}\right) = n. \tag{1}$$

   b. For any matrix $A \in \mathbb{R}^{m \times n}$

   $$\dim(\text{null}(A)) + \dim(\text{range}(A)) = n. \tag{2}$$

   c. For any matrix full-rank $A \in \mathbb{R}^{m \times n}$, where $m \leq n$, $A^T \left(AA^T\right)^{-1} Ax$ is the projection of $x \in \mathbb{R}^n$ onto the row space of A. (For *fat* matrices, which have more rows than columns, being full rank means that the rows are all linearly independent.)

   d. If the columns of $A \in \mathbb{R}^{m \times n}$, $m \geq n$, are orthonormal then the solution to the least-squares problem is of the form

   $$x = \arg\min_z ||y - Az||_2 = A^T y. \tag{3}$$

2. *Global warming (10 points).* In this problem you will implement the model described in Example of Lecture Notes 10.

   a. Complete the script *hw10_pb2.py*. Submit the code that you add (*not* the whole script, just the code you write).

   b. You show the results to a friend but he doesn't buy it and calls you a hippy for believing in global warming. You decide to do a hypothesis test to provide more convincing evidence. You fit the same model to data from 100 stations and count the number of times the slope of the linear component is positive. You choose a null hypothesis in which there is warming or cooling independently with probability 1/2 in each station. What is the p value if the slope is positive for 65 of the stations?

3. *Weight prediction (10 points).* We are interested in estimating the weight of people in a population just using their height. We would like to test two models. Model 1 is linear:

   $$\text{height} = \alpha \text{ weight}. \tag{4}$$

   Model 2 is also linear but includes an intercept

   $$\text{height} = \alpha \text{ weight} + \beta. \tag{5}$$

   a. What is the least-squares estimate of α given two training data vectors h containing n_{train} heights and w containing the corresponding weights?

b. What is the point of adding an intercept? Sketch an example of a 2D data set where this could be a good idea.

c. Complete the script *hw10_pb2.py* and report the relative errors achieved by the two models on the test dataset.

d. Try out the models using less training points (for example 100). What do you observe? What does this suggest about linear models with few parameters in terms of the number of data? Explain whether you would favor them in settings where you have a lot of data or in settings where the data is scarce and why.

4. *Noise amplification (20 points).* We are interested in estimating a vector $x \in \mathbb{R}^n$ from data $y \in \mathbb{R}^m$, $m \geq n$, that we know follows the model

$$y = Ax + z \tag{6}$$

where $A \in \mathbb{R}m \times n$ is full rank and $z \in \mathbb{R}^m$ is an unknown noise vector.

a. Write the least-squares solution x_{LS} in terms of the SVD of A, z and x.

b. Let

$$A = \begin{bmatrix} 2.1 & 1.1 \\ 3.2 & 1.6 \\ 2.4 & 1.2 \end{bmatrix}. \tag{7}$$

What is the noise vector z with unit ℓ_2 norm that maximizes the estimation error $||x_{\text{LS}} - x||_2$? What is the corresponding value of the error? (Feel free to compute the SVD using a computer.)

c. Ridge regression (also known as Tikhonov regularization in applied math) consists of optimizing the cost function

$$\min_x ||Ax - y||_2^2 + \gamma ||x||_2^2 \tag{8}$$

where $\gamma \geq 0$ is a fixed nonnegative scalar. Write (9) as a least-squares problem of the form

$$\min_x ||Bx - c||_2 . \tag{9}$$

Indicate the dimensions of B and c.

d. Show that the solution to (9) equals

$$x_{\text{RR}} := \sum_{i=1}^{n} \frac{\sigma_i}{\sigma_i^2 + \gamma^2} \left(\sigma_i v_i^T x + u_i^T z\right) v_i, \tag{10}$$

where $u_1, \ldots, u_n$ are the left singular vectors of A, $v_1, \ldots, v_n$ the right singular vectors and $\sigma_1, \ldots, \sigma_n$ the singular values. (Hint: Use the fact that $\mathrm{I} = VV^T$ where V is the matrix of right singular vectors of A)

e. Show that the error $||x_{\text{RR}} - x||_2$ can be separated into two terms, one that grows with γ and another one that decreases with γ.

f. Ridge regression is often used in cases when the matrix A is badly conditioned. Why is this a good idea?