MIDTERM
Advanced Machine Learning
DATA 442/642

Exercise 1 (10 points)
Show that the ℓ_1 norm is a convex function (as all norms), yet it is not strictly convex. In contrast, show that the squared Euclidean norm is a strictly convex function.

Exercise 2 (10 points)
Let the observations resulting from an experiment be x_n, $n = 1, 2, \ldots, N$. Assume that they are independent and that they originate from a Gaussian PDF with mean μ and standard deviation σ^2. Both, the mean and the variance, are unknown. Prove that the maximum likelihood (ML) estimates of these quantities are given by

$$\hat{\mu}_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n, \ \hat{\sigma}_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{\mu}_{ML})^2$$

Exercise 3 (15 points)
For the regression model where the noise vector $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_N]^\top$ comprises samples from zero mean Gaussian random variable, with covariance matrix $\boldsymbol{\Sigma}_n$, show that the Fisher information matrix is given by

$$I(\boldsymbol{\theta}) = \mathbf{X}^\top \boldsymbol{\Sigma}_n^{-1} \mathbf{X},$$

where $\mathbf{X}$ is the input matrix.

Exercise 4 (20 points) Consider the regression problem described in one of our labs. Read the same audio file, then add white Gaussian noise at a 15 dB level and randomly "hit" 10% of the data samples with outliers (set the outlier values to 80% of the maximum value of the data samples).

(a) Find the reconstructed data samples obtained by the support vector regression. Employ the Gaussian kernel with $\sigma = 0.004$ and set $\epsilon = 0.003$ and $C = 1$. Plot the fitted curve of the reconstructed samples together with the data used for training.

(b) Repeat step (a) using $C = 0.05, 0.1, 0.5, 5, 10, 100$.

(c) Repeat step (a) using $\epsilon = 0.0005, 0.001, 0.01, 0.05, 0.1$.

(d) Repeat step (a) using $\sigma = 0.001, 0.002, 0.01, 0.05, 0.1$.

(e) Comment on the results.

Exercise 5 (15 points)
Show, using Lagrange multipliers, that the ℓ_2 minimizer in equation (9.18) from the textbook accepts the closed form solution

$$\hat{\boldsymbol{\theta}} = \mathbf{X}^\top (\mathbf{X}\mathbf{X}^\top)^{-1} \boldsymbol{y}$$

Now, show that for the system $\boldsymbol{y} = \mathbf{X}\boldsymbol{\theta}$ with $\mathbf{X} \in \mathbb{R}^{n \times l}$ and $n > l$ the least squares solution is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{y}$$

Exercise 6 (10 points)

Show that the null space of a full rank $N \times l$ matrix $\mathbf{X}$ is a subspace of imensionality $l - N$, for $N < l$.

Exercise 7 (20 points)

Generate in Python a sparse vector $\boldsymbol{\theta} \in \mathbb{R}^l$, $l = 100$, with its first five components taking random values drawn from a normal distribution with mean zero and variance one and the rest being equal to zero. Build, also, a sensing matrix $\mathbf{X}$ with $N = 30$ rows having samples normally distributed, with mean zero and variance $1/N$, in order to get 30 observations based on the linear regression model $\boldsymbol{y} = \mathbf{X}\boldsymbol{\theta}$. Then perform the following tasks.

(a) Use a LASSO implementation to reconstruct $\boldsymbol{\theta}$ from $\boldsymbol{y}$ and $\mathbf{X}$.

(b) Repeat the experiment 500 times, with different realizations of $\mathbf{X}$, in order to compute the probability of correct reconstruction (assume the reconstruction is exact when $||\boldsymbol{y} = \mathbf{X}\boldsymbol{\theta}|| < 10^{-8}$).

(c) Repeat the same experiment (500 times) with matrices of the form

$$\mathbf{X}(i,j) = \begin{cases} +\sqrt{\frac{\sqrt{p}}{N}}, & \text{with probability } \frac{1}{2\sqrt{p}} \\ 0, & \text{with probability } 1 - \frac{1}{\sqrt{p}} \\ -\sqrt{\frac{\sqrt{p}}{N}}, & \text{with probability } \frac{1}{2\sqrt{p}} \end{cases}$$

for p equal to $1, 9, 25, 36, 64$ (make sure that each row and each column of $\mathbf{X}$ has at least a nonzero component). Give an explanation why the probability of reconstruction falls as p increases (observe that both the sensing matrix and the unknown vector are sparse).