EE227A 2/19/13
Suvrit Sra

Quiz #1

SOLUTIONS

The quiz lasts 1 hour. Notes are allowed.

Topics: Convex sets, convex functions, convex problems, Fenchel conjugates, weak duality.

1. We consider a set of m data points $x_i \in \mathbf{R}^n$, $i = 1, \dots, m$. We seek to find a line in $\mathbf{R}^n$ such that the sum of the squares of the distances from the points to the line is minimized. To simplify, we assume that the line goes through the origin.

   (a) Consider a line that goes through the origin $\mathcal{L}(v) := \{tv \, : \, t \in \mathbf{R}\}$, where $v \in \mathbf{R}^n$ is given. (You can assume without loss of generality that $\|v\|_2 = 1$.) Find an expression for the minimum Euclidean distance from a given point x to $\mathcal{L}(v)$.

   (b) Now consider the m points and find an expression for the sum of the squares of the distances from the points $x_1, \dots, x_m$ to the line $\mathcal{L}(v)$. Your expression should be of the form $c - v^T \Sigma v$, with c a constant (independent of v), and Σ a matrix which you will determine.

   (c) Consider finding a unit-norm direction v such that the sum of the squares of the distances from the points $x_1, \dots, x_m$ to the line $\mathcal{L}(v)$ is minimized. Is the problem convex ?

   (d) Find the weak dual of the problem and indicate which class the dual problem belongs to (LP, SOCP, QP, SDP).

   (e) Explain how you would find the answer to the problem in (c) (that is, find an optimal unit-norm direction v such that the sum of the squares of the distances from the points $x_1, \dots, x_m$ to the line $\mathcal{L}(v)$ is minimized), via the eigenvalue decomposition.
   Hint: Recall that the largest eigenvalue of a symmetric matrix Σ is the maximum of $v^T \Sigma v$, over the set of vectors v such that $\|v\|_2 = 1$.

Solution

(a) The projection of a single point $x \in \mathbf{R}^n$ is obtained by solving the one-dimensional least-squares problem

$$\min_t \; f(t), \quad f(t) := \|tv - x\|_2^2.$$

The optimal t is obtained by taking derivatives of the convex objective function f of this unconstrained problem. Setting $f'(t) = 0$ yields the minimizer $t^* = x^T v$. At optimum, the (squared) minimal distance is

$$f(t^*) = \|x - t^* v\|_2^2 = x^T x - 2(x^T v)t^* + (t^*)^2 = x^T x - 2(t^*)^2 + (t^*)^2 = x^T x - (x^T v)^2.$$

(b) The sum of the squares of the distances from the points $x_1, \ldots, x_m$ to the line $\mathcal{L}(v)$ is

$$D(v)^2 := \sum_{i=1}^{m} \left(x_i^T x_i - (x_i^T v)^2 \right) = c - v^T \Sigma v,$$

where

$$\Sigma := \sum_{i=1}^{m} x_i x_i^T, \quad c := \sum_{i=1}^{m} x_i^T x_i = \mathbf{Tr}\, \Sigma.$$

(c) The problem $\min_{||v||_2=1} \mathbf{Tr}\,\Sigma - v^T \Sigma v = \mathbf{Tr}\,\Sigma - \max_{||v||_2=1} v^T \Sigma v$ is non-convex since the constraint set $||v||_2 = 1$ is non-convex and objective is concave (since $\Sigma \succeq 0$, $-v^T \Sigma v$ is concave).

(d) Consider $p^* = \max_{||v||_2=1} v^T \Sigma v$. The Lagrangian is $v^T \Sigma v + \lambda(1 - v^T v)$ where λ is the dual variable. p^* has the equivalent min-max form

$$p^* = \max_{v} \min_{\lambda} v^T \Sigma v + \lambda(1 - v^T v)$$

by exchanging min and max we get the weak dual d^*:

$$\begin{aligned} p^* \leq d^* &= \min_{\lambda} \max_{v} v^T \Sigma v + \lambda(1 - v^T v) && (1) \\ &= \min_{\lambda} \max_{v} v^T (\Sigma - \lambda I) v + \lambda. && (2) \\ &= \min_{\lambda} \lambda \quad s.t. \quad \Sigma \preceq \lambda I && (3) \end{aligned}$$

Note that the dual problem is an SDP.

(e) To find v, we need to solve the problem of minimizing the quantity $D(v)^2$ by choice of v such that $\|v\|_2 = 1$. The problem reduces to

$$\max_{v \,:\, \|v\|_2 = 1} \; v^T \Sigma v.$$

As said in the hint, the optimal value of the problem is the largest eigenvalue of Σ. This value is attained by any vector v corresponding to the largest eigenvalue of Σ.

Note: We notice that $\Sigma = XX^T$, where $X := [x_1, \ldots, x_m]$ is the $n \times m$ matrix formed with the data points. If an SVD of X is known: $X = USV^T$, with S a diagonal matrix containing the singular values $\sigma_1, \ldots, \sigma_r$ of X, arranged in decreasing fashion. The solution to the above problem is $v^* = v_1$, which is a left singular vector corresponding to the largest singular value of X (that is, v^* is the first column of V).

2. (a) Show that the following function $f : \mathbf{R}^n \to \mathbf{R}$ is convex

   i. $f(x) = \lambda_{\max}(I_{10\times 10} + x_1 u_1 u_1^T + x_2 u_2 u_2^T)$, where u_1, u_2 are two given vectors in $\mathbf{R}^{10}$, $I_{10\times 10}$ is the identity matrix of size 10×10, and $\lambda_{\max}(A)$ denotes the largest eigenvalue of the symmetric matrix A.
   Hint: You can use the fact that pointwise maximum of linear (or affine) functions is convex.

   (b) Find the Fenchel conjugate $f^*(y) = \max_x \; x^T y - f(x)$ for the following functions

   i. $f(x) = \frac{1}{2} x^T A x + b^T x$ on $\mathbb{R}^n$ where $A \succ 0$

   ii. $f(x) = \max(0, 1 - x)$ on $\mathbb{R}$

Solution

(a) Due to the variational representation of eigenvalues of symmetric matrices, we have

$$f(x) = \max_{z \,:\, z^T z = 1} f_z(x),$$

where f_z is the affine function

$$f_z(x) = z^T (I_{10} + x_1 u_1 u_1^T + x_2 u_2 u_2^T) z = z^T z + x_1 (u_1^T z)^2 + x_2 (u_2^T z)^2.$$

therefore f is the pointwise maximum of affine functions hence convex.

(b) i. $f^*(y) = \max_x \; x^T y - \frac{1}{2} x^T A x - b^T x$. Equating derivatives to zero we get $Ax^* + b = y$. Since $A \succ 0$ we can invert A and write $x^* = A^{-1}(y - b)$. Plugging x^* in the objective we get, $f^*(y) = y^T A^{-1}(y - b) - \frac{1}{2}(y - b)^T A^{-1} A A^{-1}(y - b) - b^T A^{-1}(y - b) = -\frac{1}{2}(y - b)^T A^{-1}(y - b)$

ii. $f^*(y) = \max_x \; xy - \max(0, 1 - x)$.

Now we consider the first case (i) when $x \leq 1$ we have $1 - x \geq 0$ hence $\max(0, 1 - x) = 1 - x$. Then we get in case (i):

$$f^*(y) = \max_{x \leq 1} x(y+1) - 1 = \max_{x-1 \leq 0} (x-1)(y+1) + y = y \text{ if } y + 1 \geq 0, \text{ and } \infty \text{ o.w.}$$

Then we consider the second case (ii) when $x \geq 1$, we have $\max(0, 1 - x) = 0$. Then we have in case(ii),

$$f^*(y) = \max_{x \geq 1} xy = \max_{x-1 \geq 0} (x - 1)y + y = y \text{ if } y \leq 0, \text{ and } \infty \text{ o.w.}$$

Therefore,

$$f^*(y) = y \text{ if } -1 \leq y \leq 0 \text{ and } \infty \text{ o.w.}$$