

1. Introduction to optimization problem

- (a) Convex sets.
- (b) Convex functions.
- (c) Optimization problem in standard form.
 - Convex optimization.
- (d) Globally and locally optimal.
- (e) Duality.
 - Lagrange dual problem.
 - Geometric interpretation.
 - KKT conditions.

Notes: One can read the book Convex Optimization by Boyd and Vandenberghe (freely available on-line) for more extensive coverage of the above topics.

2. Find the dual problem of the following Quadratic program

$$\begin{array}{ll}\text{minimize}_x & x^T P x \\ \text{subject to} & A x \leq b\end{array}$$

Assume $P \in \mathcal{S}_{++}^n$.

Solution: The Lagrangian:

$$L(x, \lambda) = x^T P x + \lambda^T (A x - b).$$

The dual function:

$$g(\lambda) = \inf_x L(x, \lambda) = -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda.$$

The dual problem

$$\begin{array}{ll}\text{maximize}_\lambda & -\frac{1}{4} \lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \geq 0.\end{array}$$

3. Quadratic program example Consider the objective function

$$J(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2.$$

Find the optimal x that minimize $J(x)$ under the following constrains:

(a) No constrain.

Solution: Set the gradient of $J(x)$ to 0, we get:

$$10x_1 + 4x_2 + 2 = 0,$$

$$4x_1 + 4x_2 - 4 = 0.$$

The above equation give use $x = [-1, 2]^T$.

(b) $x_1 + x_2 + 2 = 0$.

Solution: We use a Lagrange multiplier λ to enforce this constrain. $L(x, \lambda) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2 + \lambda(x_1 + x_2 + 2)$. Setting the gradient with respect to x equals 0 and use this constrain, we get:

$$10x_1 + 4x_2 + 2 + \lambda = 0,$$

$$4x_1 + 4x_2 - 4 + \lambda = 0.$$

$$x_1 + x_2 + 2 = 0.$$

Solving the above equations, we get $x = [-1; -1]$ and $\lambda = 12$.

(c) $x_1 + x_2 + 2 \leq 0$.

Solution: Same as (b). See visualization.

(d) $x_1 + x_2 + 2 \geq 0$.

Solution: Same as (a). See visualization.

Visualization: See ppt for the visualization concerning this problem.

4. **Multi-class Classification Least Squares** In this section, you will determine the parameter matrix $\mathbf{W} \in \mathbb{R}^{m \times p}$ for the Multi-class Least Squares classification.

Given a data matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ and target matrix $\mathbf{T} \in \mathbb{R}^{n \times p}$, the sum-of-squares error function can be written as

$$Er(\mathbf{W}) = \text{Tr}\{(\mathbf{XW} - \mathbf{T})^T(\mathbf{XW} - \mathbf{T})\}$$

where Tr is the trace of a matrix. You can assume that \mathbf{X} has full rank.

We will solve this problem by setting the derivative with respect to \mathbf{W} to be zero and solve for \mathbf{W} . To do this we must first know some matrix derivative properties.

- (a) Let \mathbf{A}, \mathbf{Z} be two matrices. Prove

$$\frac{d\text{Tr}(\mathbf{AZ})}{d\mathbf{Z}} = \mathbf{A}^T$$

Solution:

We can do this one variable at a time. Note that

$$\text{Tr}(\mathbf{AZ}) = \sum_i \sum_j z_{ij} a_{ji}.$$

Hence, we get

$$\frac{d\text{Tr}(\mathbf{AZ})}{dz_{ij}} = a_{ji}$$

which proves the result.

- (b) Let \mathbf{A}, \mathbf{Z} be two matrices. Prove

$$\frac{d\text{Tr}(\mathbf{ZAZ}^T)}{d\mathbf{Z}} = \mathbf{ZA}^T + \mathbf{ZA}$$

Solution:

We can do this one variable at a time. Note that

$$\begin{aligned} \text{Tr}(\mathbf{ZAZ}^T) &= \sum_i \sum_j \sum_k z_{ij} a_{jk} z_{ik} \\ &= \sum_i \sum_j z_{ij}^2 a_{jk} + \sum_i \sum_j \sum_{k, k \neq j} z_{ij} a_{jk} z_{ik} \end{aligned}$$

Now, we can take the derivative with respect to each element in \mathbf{Z} .

First, let us differentiate with respect to z_{ii} and get

$$\begin{aligned} \frac{d\text{Tr}(\mathbf{ZAZ}^T)}{dz_{ii}} &= 2a_{ii}z_{ii} + \sum_{k, k \neq i} a_{ik}x_{ik} + \sum_{k, k \neq i} a_{ki}x_{ik} \\ &= \sum_k z_{ik}(a_{ik} + a_{ki}) \end{aligned}$$

which matches the result for the diagonal elements.

Now, let us differentiate with respect to z_{ij} where $i \neq j$ and get

$$\frac{d\text{Tr}(\mathbf{ZAZ}^T)}{dz_{ij}} = \sum_k z_{ik}(a_{jk} + a_{kj})$$

which matches the result for the off-diagonal elements.

- (c) Now, we can take the derivative of $Er(\mathbf{W})$ and set it to zero. Show that this results in

$$\mathbf{W} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T}$$

Solution:

$$\begin{aligned} Er(\mathbf{W}) &= \text{Tr}\{(\mathbf{XW} - \mathbf{T})^T(\mathbf{XW} - \mathbf{T})\} \\ &= \text{Tr}\{\mathbf{W}^T \mathbf{X}^T \mathbf{XW}\} - \text{Tr}\{\mathbf{W}^T \mathbf{X}^T \mathbf{T}\} - \text{Tr}\{\mathbf{T}^T \mathbf{XW}\} + \text{Tr}\{\mathbf{T}^T \mathbf{T}\} \\ \nabla_{\mathbf{W}} Er(\mathbf{W}) &= 2\mathbf{X}^T \mathbf{XW} - 2\mathbf{X}^T \mathbf{T} = 0 \\ \implies \mathbf{W} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T} \end{aligned}$$