6. HW Constrained Optimization & SVM

Dienstag, 4. Dezember 2018 11:00

1 Constrained Optimization

Problem 1: Solve the following constrained optimization problem using the recipe described in the lecture (slide 17).

$$\begin{aligned} \text{minimize} \quad & f_0(\boldsymbol{\theta}) = \theta_1 - \sqrt{3}\theta_2 \\ \text{subject to} \quad & f_1(\boldsymbol{\theta}) = \theta_1^2 + \theta_2^2 - 4 \leq 0 \end{aligned}$$

$$L(\vec{\theta}, \alpha) = \theta_1 - \sqrt{3}\theta_2 + \alpha(\theta_1^2 + \theta_2^2 - 4)$$

$$\vec{\nabla}_{\vec{\theta}} = \begin{pmatrix} 1 + 2\alpha\theta_1 \\ -\sqrt{3} + 2\alpha\theta_2 \end{pmatrix} = 0$$

$$\Rightarrow \theta_1 = -\frac{1}{2\alpha}, \quad \theta_2 = \frac{\sqrt{3}}{2\alpha}$$

$$g(\alpha) = \frac{-1}{2\alpha} - \frac{3}{2\alpha} + \frac{1}{4\alpha} + \frac{3}{4\alpha} - 4\alpha = -\frac{1}{\alpha} - 4\alpha$$

$$\partial_\alpha g(\alpha) = \frac{1}{\alpha^2} - 4 = 0$$

$$\alpha^2 = \frac{1}{4}$$

$$\alpha_{1,2} = \pm\frac{1}{2}$$

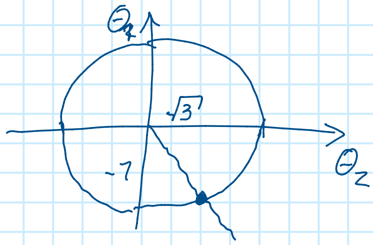
$$\alpha^* = \frac{1}{2}$$

$$p^* = g\left(\frac{1}{2}\right) = -4$$

$$\theta_1^* = -1, \quad \theta_2^* = \sqrt{3}$$

$$\theta_1^2 + \theta_2^2 \leq 4$$

$$\|\vec{\theta}\|_2^2 \leq 4 \Leftrightarrow \|\vec{\theta}\|_2 \leq 2$$

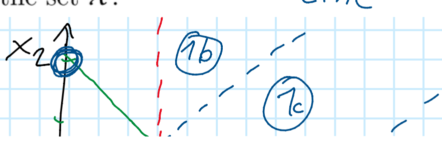


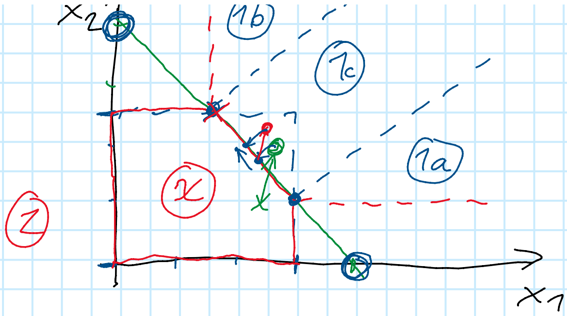
2 Projected Gradient Descent

Problem 2: Given is the following (convex) domain defined by a set of linear constraints

$$\mathcal{X} \subset \mathbb{R}^2 = \{\boldsymbol{x} \in \mathbb{R}^2 : \underbrace{(x_1 + x_2 \leq 4)}_{\text{Line}} \wedge \underbrace{(0 \leq x_1 \leq 3) \wedge (0 \leq x_2 \leq 2.5)}_{\text{box constraint}}\}.$$

a) Visualize the set $\mathcal{X}$.





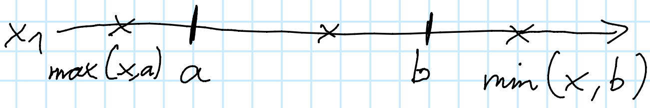
b) Derive a closed form for the projection $\pi_{\mathcal{X}}(\boldsymbol{p}) = \arg\min_{\boldsymbol{x}\in\mathcal{X}} ||\boldsymbol{x}-\boldsymbol{p}||_2^2$. That is, given an arbitrary point $\boldsymbol{p} \in \mathbb{R}^2$, what is its projection on $\mathcal{X}$?

Hint: For one part of $\mathbb{R}^2$ you might want to use the line projection $\pi_{\text{line}}(\boldsymbol{p}) = \boldsymbol{a} + \frac{(\boldsymbol{p}-\boldsymbol{a})^T(\boldsymbol{b}-\boldsymbol{a})}{||\boldsymbol{b}-\boldsymbol{a}||_2^2}(\boldsymbol{b}-\boldsymbol{a})$, where $\boldsymbol{a} \in \mathbb{R}^2$ and $\boldsymbol{b} \in \mathbb{R}^2$ specify the line.

$$\pi_{\mathcal{X}}(\vec{p}) = \begin{cases} \vec{p} & \text{if } \vec{p} \in \mathcal{X} \\ (3,1)^T & \text{if } \vec{p} \in \{\vec{p} \mid p_2 \geq 1 \wedge -p_1+p_2 \leq -2\} \quad 1a \\ (1{,}5;2{,}5)^T & \text{if } \vec{p} \in \{\vec{p} \mid p_1 \geq 1{,}5 \wedge -p_1+p_2 \geq 1\} \\ \pi_{line}(\vec{p}) & \text{if } \vec{p} \in \{\vec{p} \mid -2 < -p_1+p_2 < 1 \wedge p_1+p_2 > 4\} \\ \pi_{box}(\vec{p}) & \text{if } \vec{p} \in \{\vec{p} \mid \vec{p} \notin \mathcal{X} \wedge (p_1 < 1{,}5 \vee p_2 < 1)\} \end{cases}$$

$$\pi_{line}(\vec{p}) = \vec{a} \, \frac{(\vec{p}-\vec{a})^T(\vec{b}-\vec{a})}{||\vec{b}-\vec{a}||_2^2}(\vec{b}-\vec{a}) \overbrace{= \ldots =}^{\vec{a}=\binom{4}{0},\ \vec{b}=\binom{0}{4}} \begin{pmatrix} 2+\frac{1}{2}p_1-\frac{1}{2}p_2 \\ 2-\frac{1}{2}p_1+\frac{1}{2}p_2 \end{pmatrix}$$

$$\pi_{box}(\vec{p}) = \begin{pmatrix} \max(0, \min(3, p_1)) \\ \max(0, \min(2{,}5, p_2)) \end{pmatrix}$$



c) Given is the following constrained optimization problem:

$$\min_{\boldsymbol{x}} \; (x_1-2)^2 + (2x_2-7)^2,$$
$$\text{subject to } \boldsymbol{x} \in \mathcal{X}.$$

Perform two steps of projected gradient descent starting from the point $\boldsymbol{x}^{(0)} = (2.5, 1)^T$. Use a constant learning rate/step size of $\tau = 0.05$.

1. $\vec{p}^{\,t+1} = \vec{x}^{\,t} - \tau \vec{\nabla}_{\vec{x}} f(\vec{x}^{\,t})$
2. $\vec{x}^{\,t+1} = \pi_{\mathcal{X}}(\vec{p}^{\,t+1})$

$$2.\ \vec{x}^{\,t+1} = \pi_{\mathcal{X}}(\vec{p}^{\,t+1})$$

$$\vec{\nabla}_{\vec{x}} f(\vec{x}) = \begin{pmatrix} 2x_1 - 4 \\ 8x_2 - 28 \end{pmatrix}$$

1.1 $\vec{p}^{(1)} = \vec{x}^{(0)} - \tau \vec{\nabla}_{\vec{x}} f(\vec{x}^{(0)}) = \begin{pmatrix} 2{,}45 \\ 2 \end{pmatrix}$

1.2 $\vec{x}^{(1)} = \pi_{\mathcal{X}}(\vec{p}^{(1)}) = \begin{pmatrix} 2{,}225 \\ 1{,}775 \end{pmatrix}$

2.1 $\vec{p}^{(2)} = \begin{pmatrix} 2{,}2025 \\ 2{,}465 \end{pmatrix}$

2.2 $\vec{x}^{(2)} = \begin{pmatrix} 1{,}86875 \\ 2{,}13125 \end{pmatrix}$

3 SVM

Problem 3: Explain the similarities and differences between the SVM and perceptron algorithms.

Sim.: Both look for a hyperplane that separates 2 classes

Diff.: SVM maximizes margin

Problem 4: Show that the duality gap is zero for SVM.

Slater's condition: 1. Convex objective ✓

2. Constraint affine ✓

$\exists$ solution inside constraints

Example: $f_1(\vec{x}) = x_1^2 + x_2^2 - 4 \leq 0$

<

$x_1 = 0,\ x_2 = 0$: $\quad 0^2 + 0^2 - 4 = -4 < 0$

Problem 5: Recall that the dual function for SVM (slide 41) can be written as

$$g(\boldsymbol{\alpha}) = \boxed{\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{Q} \boldsymbol{\alpha}} + \boxed{\boldsymbol{\alpha}^T \mathbf{1}_N}$$

(a) Show how the matrix $\boldsymbol{Q}$ can be computed. *(Hint: You might want to use Hadamard product, denoted as $\odot$).*

$$g(\vec{\alpha}) = \boxed{\sum_i \alpha_i} \boxed{-\frac{1}{2}\sum_i \sum_j y_i y_j \alpha_i \alpha_j \vec{x}_i^T \vec{x}_j}$$

$$-\frac{1}{2}\sum_i \sum_j \alpha_i\, y_i y_j \underbrace{\vec{x}_i^T \vec{x}_j}\, \alpha_j =$$

$$= -\frac{1}{2}\sum_i \sum_j \alpha_i\, y_i y_j \sum_k X_{ik} \overbrace{X_{jk}}^{(X^T)_{kj}} \alpha_j =$$

$$= -\frac{1}{2}\sum_i\sum_j \alpha_i \underbrace{y_i y_j}_{Y_{ij}=y_iy_j} \underbrace{\sum_k X_{ik} \overbrace{X_{jk}}^{(X^T)_{kj}}}_{X \cdot X^T} \alpha_j =$$

$$Y = \vec{y}\vec{y}^T$$

$$= -\frac{1}{2}\sum_i\sum_j \alpha_i \underbrace{\underbrace{Y_{ij}(XX^T)_{ij}}_{(Y \odot XX^T)_{ij}}}_{-Q_{ij}} \alpha_j$$

$$Q = -(\vec{y}\vec{y}^T \odot XX^T)$$

(b) Prove that the matrix $\boldsymbol{Q}$ is negative (semi-)definite.

$\Leftrightarrow$ $M = \vec{y}\vec{y}^T \odot XX^T$ is pos. (semi-)def.

$\forall \vec{a}$ $\quad \vec{a}^T M \vec{a} \geq 0$

$$\vec{a}^T M \vec{a} = \sum_{i=1}^N \sum_{j=1}^N a_i y_i y_j \overbrace{\vec{x}_i^T \vec{x}_j}^{XX^T} a_j =$$

$$= \underbrace{(\vec{a} \odot \vec{y})^T X}\, X^T (\vec{a} \odot \vec{y}) =$$

$$= (X^T(\vec{a} \odot \vec{y}))^T (X^T(\vec{a} \odot \vec{y})) =$$

$$= (X^T(\vec{a} \odot \vec{y}))^2 \geq 0$$

$\Leftrightarrow$ $Q = -M$ is neg. (semi-) def. $\square$

(c) Explain what the negative (semi-)definiteness means for our optimization problem. Why is this property important?

$\Rightarrow$ Maximization problem is concave

$\Rightarrow$ local max is global max.

Problem 6: Download the notebook `homework_06_notebook.ipynb` from Piazza. Fill in the missing code and run the notebook. Convert the evaluated notebook to pdf and add it to the printout of your homework (see printing instructions inside the notebook).

Quad. opt.: $\min_{\vec{x}} \quad \frac{1}{2}\vec{x}^T P \vec{x} + \vec{q}^T \vec{x}$

subject to $\quad G\vec{x} \leq \vec{h}$

$A\vec{x} = \vec{b}$

SVM: Dual $\underline{\max}$ $\quad g(\vec{\alpha}) = \frac{1}{2}\vec{\alpha}^T Q \vec{\alpha} + \vec{\alpha}^T \vec{1}_N$

$\vec{y}^T\vec{\alpha} = 0$

$\alpha_i \geq 0 \quad \forall i$

$$\vec{x} \mathrel{\hat{=}} \vec{\alpha}$$

$$P \mathrel{\hat{=}} -Q = \sum_i \sum_j y_i y_j \, \vec{x}_i^T \vec{x}_j = \sum_i \sum_j y_i y_j \sum_k X_{ik} X_{jk} =$$

np.einsum("i, j, ik, jk -> ij", $\vec{y}$, $\vec{y}$, X, X)

$$= \sum_i \sum_j \sum_k (y_i X_{ik})(y_j X_{jk}) = (Y \odot X)(Y \odot X)^T$$

$$Y = (\vec{y}, \vec{y}, \vec{y}, \dots)$$

$$\vec{q} \mathrel{\hat{=}} -\vec{1}$$
$$G \mathrel{\hat{=}} -I_N$$
$$\vec{h} \mathrel{\hat{=}} 0$$
$$A \mathrel{\hat{=}} \vec{y}^T$$
$$b \mathrel{\hat{=}} 0$$

2) $w_j = \sum_i \alpha_i y_i X_{ij} = [X^T(\vec{\alpha} \odot \vec{y})]_j$

np.einsum("i, i, ij -> j", $\vec{\alpha}$, $\vec{y}$, X)

$b_{temp,i} = y_i - \sum_j X_{ij} w_j \qquad i \in S$: set of support vectors $(\alpha_i > 0)$

$b = \frac{1}{|S|} \sum_{i \in S} b_{temp,i}$ more stable