Machine Learning 2024 - Sheet 3.1
Block III: Optimization

Isabel Valera

Exercise 1: Quadratic Over Linear Function ⚡

Consider the quadratic-over-linear function given in equation 1.

i) Show the condition for which the function is convex

$$f(x,y) = \frac{x^2}{y} \tag{1}$$

Exercise 2: Lagrange Dual Problem ⚡⚡

1. Consider the optimization problem in Eq. 2 with variable $x \in \mathbb{R}$ given in equation.
   (a) Formulate the dual problem.
   (b) What is the optimal solution for the dual or primal problem?
   (c) Verify that the dual formulation is a convex minimization problem.

$$\begin{aligned} &\text{maximize } x + y \\ &\text{subject to } x^2 + y^2 = 1 \end{aligned} \tag{2}$$

Exercise 3: Inequality constraint ⚡⚡⚡⚡

i) Express the dual problem of the primal problem given in 3 with $c \neq 0$ in terms of the conjugate f^*.

ii) Explain why the dual problem you give is convex. We do not assume f is convex.

$$\begin{aligned} &\text{minimize } c^\top x \\ &\text{subject to } f(x) \leq 0 \end{aligned} \tag{3}$$

Hint: The conjugate of a function $f^*(y) = \sup_{x \in \mathbf{dom} f} (y^T x - f(x))$

Hint: The perspective of a function $g(x,t) = tf(x/t)$

Exercise 4: KKT conditions ⚡⚡⚡

i) Derive the KKT conditions for the problem given in problem 4 with variable $X \in \mathbf{S}^n$ (n-dimensional symmetric) and domain $\mathbf{S}^n_{++}$ (symmetric positive-definite). $y \in \mathbb{R}^n$ and $s \in \mathbb{R}^n$ are given with $s^T y = 1$.

ii) Verify that the optimal solution is given by equation 5.

$$\begin{aligned} &\text{minimize } \mathbf{tr} X - \log\det X \\ &\text{subject to } Xs = y \end{aligned} \tag{4}$$

$$X^* = I + yy^T - \frac{1}{s^T s} ss^T \tag{5}$$

Exercise 5: Estimating covariance and mean ⚡⚡

Exercise 7.4 in Boyd's book

We consider the problem of estimating the covariance matrix $\mathbf{\Sigma}$ and the mean $\boldsymbol{\mu}$ of a Gaussian probability density function as given in equation 6 based on N independent samples $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N \in \mathbb{R}^n$.

i) We first consider the estimation problem when there are no additional constrains on $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\mathbf{\Sigma}}$ be the sample mean and covariance as defined in equations 7. Show that the log-likelihood function given in equation 8 can be expressed as in equation 9.

ii) Use this expression to show that if $\hat{\mathbf{\Sigma}} \succ 0$ the ML estimates of $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$ are unique and given by the sample covariance and sample mean.

$$p(\boldsymbol{x} \mid \mathbf{\Sigma}, \boldsymbol{\mu}) = (2\pi)^{-\frac{n}{2}} \det(\mathbf{\Sigma})^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{6}$$

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{x}_k \\ \hat{\mathbf{\Sigma}} &= \frac{1}{N}\sum_{k=1}^{N} (\boldsymbol{x}-\hat{\boldsymbol{\mu}})(\boldsymbol{x}-\hat{\boldsymbol{\mu}})^\top \end{aligned} \tag{7}$$

$$\mathcal{L}(\mathbf{\Sigma}, \boldsymbol{\mu}) = -\frac{Nn}{2}\log(2\pi) - \frac{N}{2}\log\det\mathbf{\Sigma} - \frac{1}{2}\sum_{k=1}^{N}(\boldsymbol{x}_k-\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1}(\boldsymbol{x}_k-\boldsymbol{\mu}) \tag{8}$$

$$\mathcal{L}(\mathbf{\Sigma}, \boldsymbol{\mu}) = \frac{N}{2}\left(-n\log(2\pi) - \log\det\mathbf{\Sigma} - \mathbf{tr}(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}) - (\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})^\top \mathbf{\Sigma}^{-1}(\boldsymbol{\mu}-\hat{\boldsymbol{\mu}})\right) \tag{9}$$

Exercise 6: Estimating mean and variance ⚡⚡

Consider a random variable $x \in \mathbb{R}$ with density p, which is normalized, i.e. has zero mean and unit variance. Consider a random variable $y = \frac{x+b}{a}$ obtained by an affine transformation of x, where $a > 0$. The random variable y has mean $\frac{b}{a}$ and variance $\frac{1}{a^2}$. As a and b vary over the non-negative real numbers $\mathbb{R}_+$ and the real numbers $\mathbb{R}$, respectively, we generate a family of densities obtained from p by scaling and shifting, uniquely parametrized by mean and variance.

i) Show that if p is log-concave, then finding the ML estimate of a and b, given samples $y_1, ..., y_n$ of y is a convex problem.

ii) As an example, work out an analytical solution for the ML estimates of a and b, assuming p is a normalized Laplacian density $p(x) = \exp(-2|x|)$.

Hint: Apply the rule of density transformation (change of variables) to obtain p_y.

Exercise 7: Robust linear classification ⚡⚡⚡

Consider the robust linear classification problem given in problem 10 where we seek an affine function $f(x) = \mathbf{w}^\top \boldsymbol{x} - b$ that separates the two sets of points $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$ and $\{\boldsymbol{y}_1, ..., \boldsymbol{y}_M\}$. This means that $\mathbf{w}^\top \boldsymbol{x}_i - b > 0$ for $i = 1, ..., N$ and $\mathbf{w}^\top \boldsymbol{y}_j - b < 0$ for $j = 1, ..., M$.

i) Show that the optimal value t^* is positive if and only if the two sets of points can be linearly separated.

ii) When the two sets of points can be linearly separated, show that the inequality $||\mathbf{w}||_2 \leq 1$ is tight, i.e., we have $||\mathbf{w}^*||_2 = 1$ for the optimal $\mathbf{w}^*$.

iii) Using the change of variables $\tilde{\mathbf{w}} = \frac{\mathbf{w}}{t}, \tilde{b} = \frac{b}{t}$, prove that problem 10 is equivalent to the quadratic program given in 11.

$$
\begin{aligned}
&\text{maximize } t \\
&\text{subject to } \mathbf{w}^\top \boldsymbol{x}_i - b \geq t,\ i = 1, ..., N \\
&\qquad\qquad\ \mathbf{w}^\top \boldsymbol{y}_i - b \leq -t,\ i = 1, ..., M \\
&\qquad\qquad\ ||\mathbf{w}||_2 \leq 1
\end{aligned} \tag{10}
$$

$$
\begin{aligned}
&\text{minimize } ||\tilde{\mathbf{w}}||_2 \\
&\text{subject to } \tilde{\mathbf{w}}^\top \boldsymbol{x}_i - \tilde{b} \geq 1,\ i = 1, ..., N \\
&\qquad\qquad\ \tilde{\mathbf{w}}^\top \boldsymbol{y}_i - \tilde{b} \leq -1,\ i = 1, ..., M
\end{aligned} \tag{11}
$$

References

[1] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge university press, 2004.