Optimization

Multivariate Roots

Learning goals

- LEARNING GOAL 1
- LEARNING GOAL 2

CONSTRAINED OPTIMIZATION IN STATISTICS

Example: Maximum Likelihood Estimation

For data $\left(\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n)}\right)$, we want to find the maximum likelihood estimate

$$\max_{\theta} L(\theta) = \prod_{i=1}^{n} f(^{(i)}, \theta)$$

In some cases, θ can only take **certain values.**

- If f is a Poisson distribution, we require the rate λ to be non-negative, i.e. $\lambda \geq 0$

CONSTRAINED OPTIMIZATION IN STATISTICS

- If f is a multinomial distribution

$$f(x_1, ..., x_p; n; \theta_1, ..., \theta_p) = \begin{cases} \binom{n!}{x_1! \cdot x_2! \ldots x_p!} \theta_1^{x_1} \cdot ... \cdot \theta_p^{x_p} & \text{if } x_1 + ... + x_p = n \\ 0 & \text{else} \end{cases}$$

  The probabilities θ_i must lie between 0 and 1 and add up to 1, i.e. we require

$$0 \leq \theta_i \leq 1 \qquad \text{for all } i$$
$$\theta_1 + ... + \theta_p = 1.$$

CONSTRAINED OPTIMIZATION IN ML

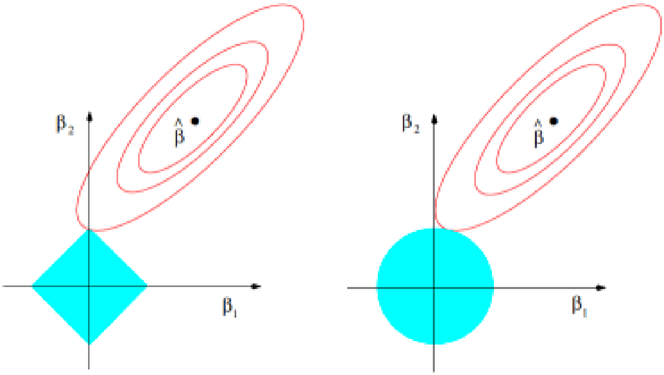
- **Lasso regression**:

$$\begin{aligned} \min_{\boldsymbol{\beta}} \quad & \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{y}^{(i)} - \boldsymbol{\theta}^T \mathbf{x}^{(i)}\right)^2 \\ \text{s.t.} \quad & \|\boldsymbol{\theta}\|_1 \leq t \end{aligned}$$

- **Ridge regression**:

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{y}^{(i)} - \boldsymbol{\theta}^T \mathbf{x}^{(i)}\right)^2 \\ \text{s.t.} \quad & \|\boldsymbol{\theta}\|_2 \leq t \end{aligned}$$

CONSTRAINED OPTIMIZATION IN ML



CONSTRAINED OPTIMIZATION IN ML

Remember the dual formulation of the SVM, which is a convex quadratic program with box constraints plus one linear constraint:

$$
\begin{aligned}
\max_{\boldsymbol{\alpha}] \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left\langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \right\rangle \\
\text{s.t.} \quad & 0 \leq \alpha_i \leq C, \\
& \sum_{i=1}^{n} \alpha_i y^{(i)} = 0,
\end{aligned}
$$

CONSTRAINED OPTIMIZATION

General definition of a **Constrained Optimization problem**:

$$\begin{aligned} \min \quad & f(\mathbf{x}) & \\ \text{such that} \quad & g_i(\mathbf{x}) \leq 0 & \text{for } i = 1, \ldots, k \\ & h_j(\mathbf{x}) = 0 & \text{for } j = 1, \ldots, l, \end{aligned}$$

where

- $g_i : \mathbb{R}^d \to \mathbb{R}, i = 1, ..., k$ are inequality constraints,
- $h_j : \mathbb{R}^d \to \mathbb{R}, j = 1, ..., l$ are equality constraints.

The set of inputs **x** that fulfill the constraints, i.e.

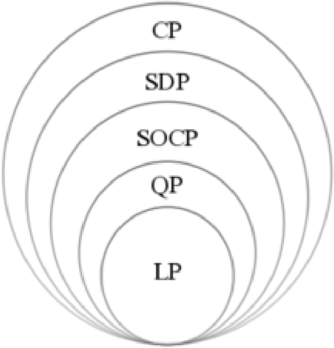
$$\mathcal{S} := \{\mathbf{x} \in \mathbb{R}^d \mid g_i(\mathbf{x}) \leq 0, h_j(\mathbf{x}) = 0 \ \forall\ i, j\}$$

is known as the feasible set.

CONSTRAINED OPTIMIZATION

Special cases of constrained optimization problems are **convex programs**, with convex objective function f, convex inequality constraints g_i, and affine equality constraints h_j (i.e. $h_j(\mathbf{x}) = \boldsymbol{a}_j^\top \mathbf{x} - \boldsymbol{b}_j$).

Convex programs can be categorized into



CONSTRAINED OPTIMIZATION

- Linear program (LP): objective function f and all constraints g_i, h_j are linear functions
- Quadratic program (QP): objective function f is a quadratic form, i.e.

  $$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \boldsymbol{Q}\mathbf{x} + \boldsymbol{c}^\top \mathbf{x} + \boldsymbol{d}$$

  for $\boldsymbol{Q} \in \mathbb{R}^{d\times d}, \boldsymbol{c} \in \mathbb{R}^d, d \in \mathbb{R}$, and constraints are linear.

as well as second-order cone programs (SOCP), semidefinite programs (SDP), cone programs (CP), and graph form programs (GFP).