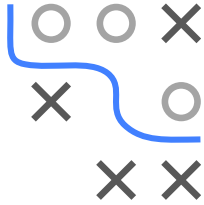
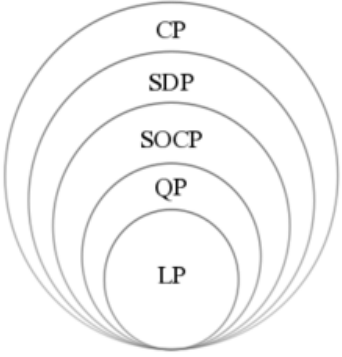
Optimization in Machine Learning



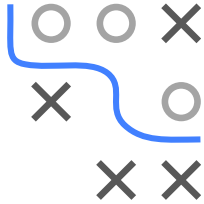
Nonlinear programs
Solvers



Learning goals

- Definition
- Max. Likelihood
- Normal regression
- Risk Minimization

SEQUENTIAL QUADRATIC PROGRAMMING



For simplification, we consider only equality constraints, thus problems of the form

$$\min f(\mathbf{x}) \qquad \text{s.t.} \qquad h(\mathbf{x}) = 0.$$

Idea:

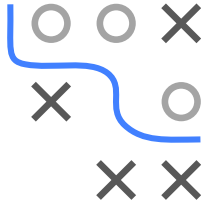
- Instead of f we optimize the 2nd order Taylor approximation in a point $\tilde{\boldsymbol{x}}$

$$\tilde{f}(\mathbf{x}) = f(\tilde{\boldsymbol{x}}) + \nabla_x f(\tilde{\boldsymbol{x}})^T(\mathbf{x} - \tilde{\boldsymbol{x}}) + \frac{1}{2}(\mathbf{x} - \tilde{\boldsymbol{x}})^T \nabla^2_{xx} f(\tilde{\boldsymbol{x}})(\mathbf{x} - \tilde{\boldsymbol{x}})$$

- h is also replaced by its linear approximation in $\tilde{\boldsymbol{x}}$.

$$\tilde{h}(\mathbf{x}) = h(\tilde{\boldsymbol{x}}) + \nabla h(\tilde{\boldsymbol{x}})^T(\mathbf{x} - \tilde{\boldsymbol{x}}).$$

PENALTY METHODS



Idea: Replace the constrained Optimization problem with a sequence of unconstrained optimization problems using a **penalty function**.

Instead of looking at

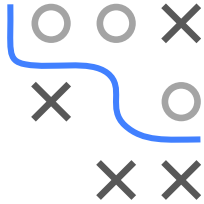
$$\min f(\mathbf{x}) \qquad \text{s.t.} \quad h(\mathbf{x}) = 0.$$

we look at the unconstrained optimization problem

$$\min_{\mathbf{x}} p(\mathbf{x}) = f(\mathbf{x}) + \rho \frac{\|h(\mathbf{x})\|^2}{2}.$$

Under appropriate conditions it can be shown that the solutions of the problem for $\rho \to \infty$ converge against the solution of the initial problem.

BARRIER METHOD



Idea: Establish a “barrier” that penalizes if **x** comes too close to the edge of the allowed set ***S***. For the problem

$$\min f(\mathbf{x}) \qquad \text{s.t.} \quad g(\mathbf{x}) \leq 0$$

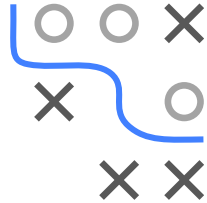
a common **Barrier function** is

$$B_\rho = f(\mathbf{x}) - \rho \sum_{i=1}^{m} \ln(-g_i(\mathbf{x}))$$

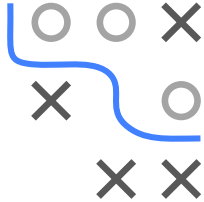
The penalty term becomes larger, the closer **x** comes to 0, i.e. the limit of the feasible set. Under certain conditions, the solutions of $\min B_\rho$ for $\rho \to 0$ converge against the optimum of the original problem.

The procedure is also called **interior-point method**.

Constrained Optimization in R



CONSTRAINED OPTIMIZATION IN R



- The function **optim(..., method = "L-BFGS-B")** uses quasi-newton methods and can handle box constraints.
- The function **nlminb()** uses trust-region procedures and can also handle box constraints.
- **constrOptim()** can be used for optimization problems with linear inequality conditions and is based on interior-point methods.
- **nloptr** is an interface to **NLopt**, an open-source library for nonlinear optimization (`https://nlopt.readthedocs.io/en/latest/`)