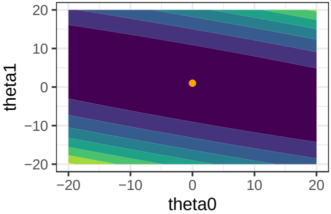
Optimization

Smooth unconstrained problems



Learning goals

- TODO
- TODO

GENERAL DEFINITION

Consider the **optimization problem**

$$\min_{\mathbf{x} \in \mathcal{S} \subseteq \mathbb{R}^d} f(\mathbf{x})$$

with objective function

$$f: \mathcal{S} \rightarrow \mathbb{R}.$$

The problem is called

- **unconstrained**, if the domain $\mathcal{S}$ is not restricted:

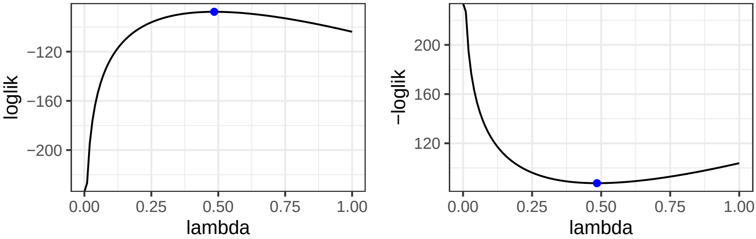
$$\mathcal{S} = \mathbb{R}^d$$

- **smooth** if f is smooth.
- **univariate** if $d = 1$, and **multivariate** if $d > 1$.

NOTE: A CONVENTION IN OPTIMIZATION

W.l.o.g., we always **minimize** functions f.

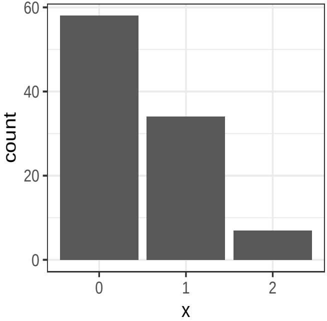
Maximization results from minimizing $-f$.



Poisson example: Maximizing the log-likelihood (left) is equivalent to minimizing the negative log-likelihood (right).

EXAMPLE 1: MAXIMUM LIKELIHOOD ESTIMATION

Assume an i.i.d. sample $\mathcal{D} = \left(x^{(1)}, ..., x^{(n)}\right)$ from a distribution with density $f(x \mid \theta)$. We want to find λ which makes the observed data most likely.



Example: Histogram of a sample drawn from a Poisson distribution

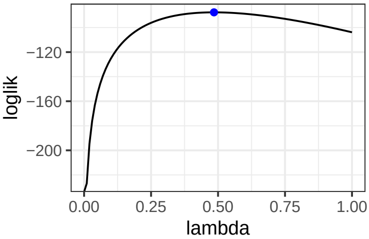
$$f(k \mid \lambda) := \mathbb{P}(x = k) = \frac{\lambda^k \cdot \exp(-\lambda)}{k!}.$$

EXAMPLE 1: MAXIMUM LIKELIHOOD ESTIMATION

We operationalize this as **maximizing** the log-likelihood function (or equivalently: minimizing the negative log-likelihood) with respect to λ:

$$
\begin{aligned}
\hat{\lambda} &= \text{arg min}_{\lambda} \; -\ell(\lambda, \mathcal{D}) = \text{arg min}_{\lambda} - \log \mathcal{L}(\lambda, \mathcal{D}) = \text{arg min}_{\lambda} - \log \prod_{i=1}^{n} f\left(\mathbf{x}^{(i)} \mid \lambda\right) \\
&= \text{arg min}_{\lambda} - \sum_{i=1}^{n} f\left(x^{(i)} \mid \lambda\right) = \text{arg min}_{\lambda} \sum_{i=1}^{n} \frac{-\lambda^{\mathbf{x}^{(i)}} \cdot \exp(-\lambda)}{\mathbf{x}^{(i)}!}
\end{aligned}
$$

EXAMPLE 1: MAXIMUM LIKELIHOOD ESTIMATION

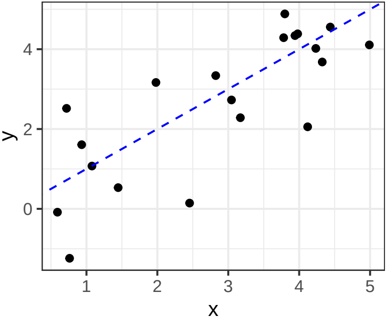


Example: The log-likelihood of a Poisson distribution for data example above. The objective function is univariate and differentiable, and the domain is unconstrained.

EXAMPLE 2: NORMAL REGRESSION

Assume a dataset $\mathcal{D} = \left(\left(\mathbf{x}^{(1)}, y^{(1)}\right), \ldots, \left(\mathbf{x}^{(n)}, y^{(n)}\right)\right)$ generated according to

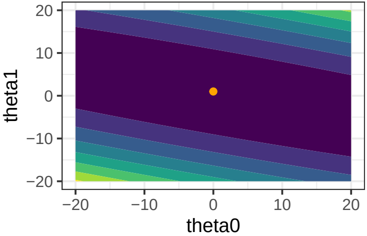
$$y^{(i)} = \boldsymbol{\theta}^\top \mathbf{x}^{(i)} + \epsilon^{(i)}, \qquad \epsilon^{(i)} \overset{iid}{\sim} \mathcal{N}\left(0, 1\right).$$



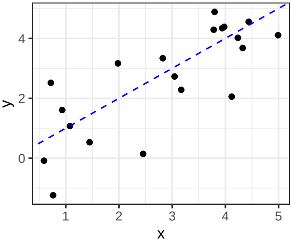
EXAMPLE 2: NORMAL LINEAR REGRESSION

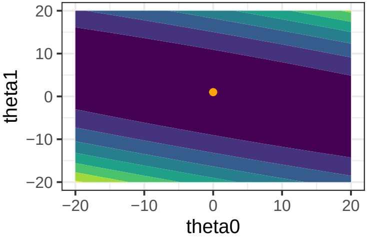
In normal linear regression the goal is to find a vector $\boldsymbol{\theta}$ which minimizes the sum of squared errors (SSE):

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^d} \sum_{i=1}^{n} \left(\boldsymbol{\theta}^\top \mathbf{x}^{(i)} - y^{(i)} \right)^2$$



EXAMPLE 2: NORMAL REGRESSION





- The problem is multivariate, smooth, and unconstrained
- Since the problem is a quadratic form, we easily obtain a geometric interpretation of the problem
- The problem has a closed-form solution, which is given by $\boldsymbol{\theta} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{y}$, where $\mathbf{X}$ is the design matrix