Friedrich-Alexander-Universität
Technische Fakultät

Machine Learning
Data Analytics

Bayesian Linear Regression

Richard Dirauf, M.Sc.
Machine Learning and Data Analytics (MaD) Lab
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
MLTS Exercise, 16.11.2022

Announcement

MLTS Exercise 02

Exam:

- Thursday, 23. March 2023
- 9:00 – 10:30
- Online exam

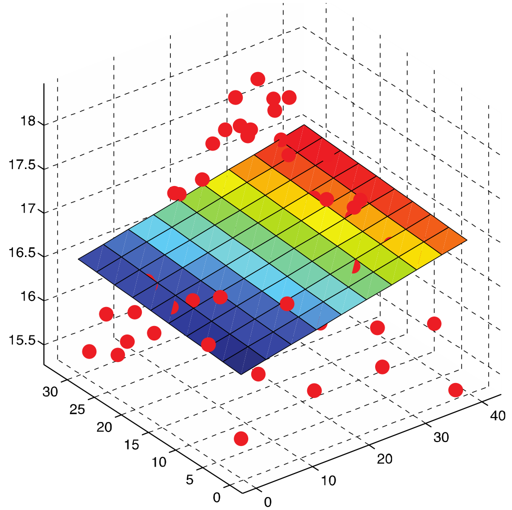
What is Linear Regression?

MLTS Exercise 02

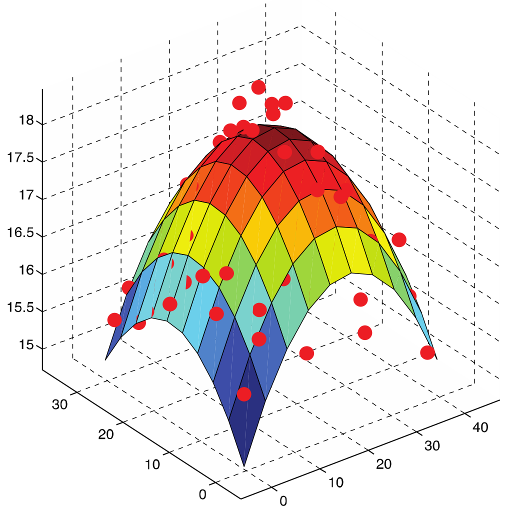
Given some tuples in a dataset:

$$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$$

We want to predict a scalar $\boldsymbol{y}$ response with one or multiple explanatory variables $\boldsymbol{x}$



$$y = w_0 + w_1 x_1 + w_2 x_2$$



$$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$$

?

What is Linear Regression?

MLTS Exercise 02

Given: $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

Where: $X \in \mathcal{R}^D, \ y \in \mathcal{R}$

Find: $f_w : \mathcal{R}^D \to \mathcal{R}$

Predict:

$$y = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_N x_N + \epsilon$$

Noise Error e.g.: $\epsilon = \mathcal{N}(\mathbf{0}, \mathbf{1})$

→ Random sampling noise or effect of variables not included in the model

Ordinary Least Squares (Smallest Residual Error)

$$RSS(w) = \sum_{i=1}^{M} (y_i - w^T x_i)^2$$

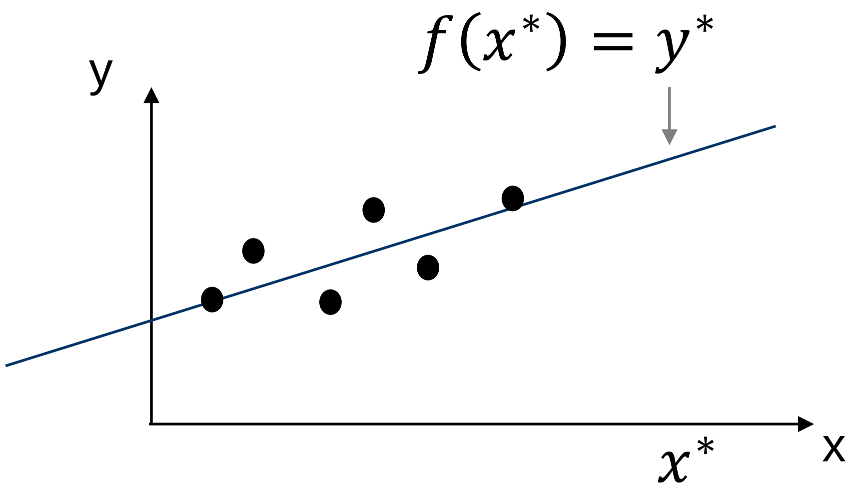
Find parameters:

$$w^* = (X^T X)^{-1} X^T y$$

Predict:

$$y^* = f(x^*) = w^T x^* = \sum_{i=1}^{d} w_i {x^*}^{(i)}$$

We only get a point estimate!



What to do instead?

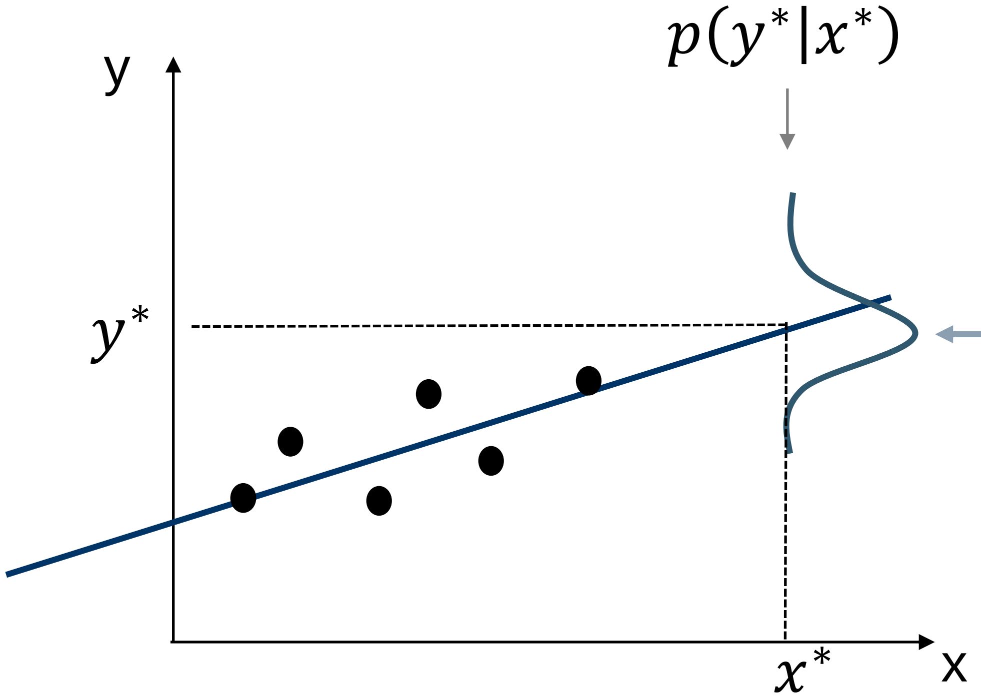
Get **distribution** of possible **y** values given **x**

$$p(y|x)$$

Formulate LR using probability distributions instead of point estimates:

$$p(y|x) = \mathcal{N}(y|\mu(x), \sigma^2(x));\ \theta = (\mu, \sigma^2)$$

Get **distribution** of possible **y** values given **x**



Some distribution (e.g., Gaussian)

Bayes Rule:

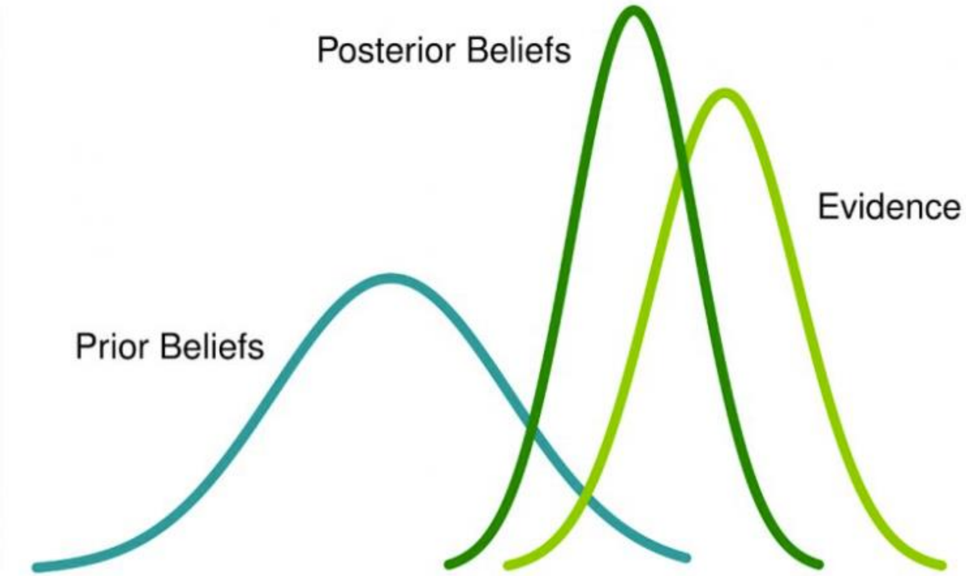
Likelihood

Prior

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Posterior

Marginal Likelihood
(Evidence)

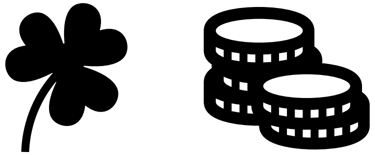


?

Example

MLTS Exercise 02

What is the probability of the outcome of a coin flip game being fair?



$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}$$

Observed data $\boldsymbol{\mathcal{D}}$

Model Parameters $\boldsymbol{\theta}$

Evidence $\boldsymbol{p(\mathcal{D})}$ → Probability of observing data across all possible θ

Prior $\boldsymbol{p(\theta)}$ → Believe of the fairness of the coin $\mathrm{p}(\theta) \in [0, 1]$

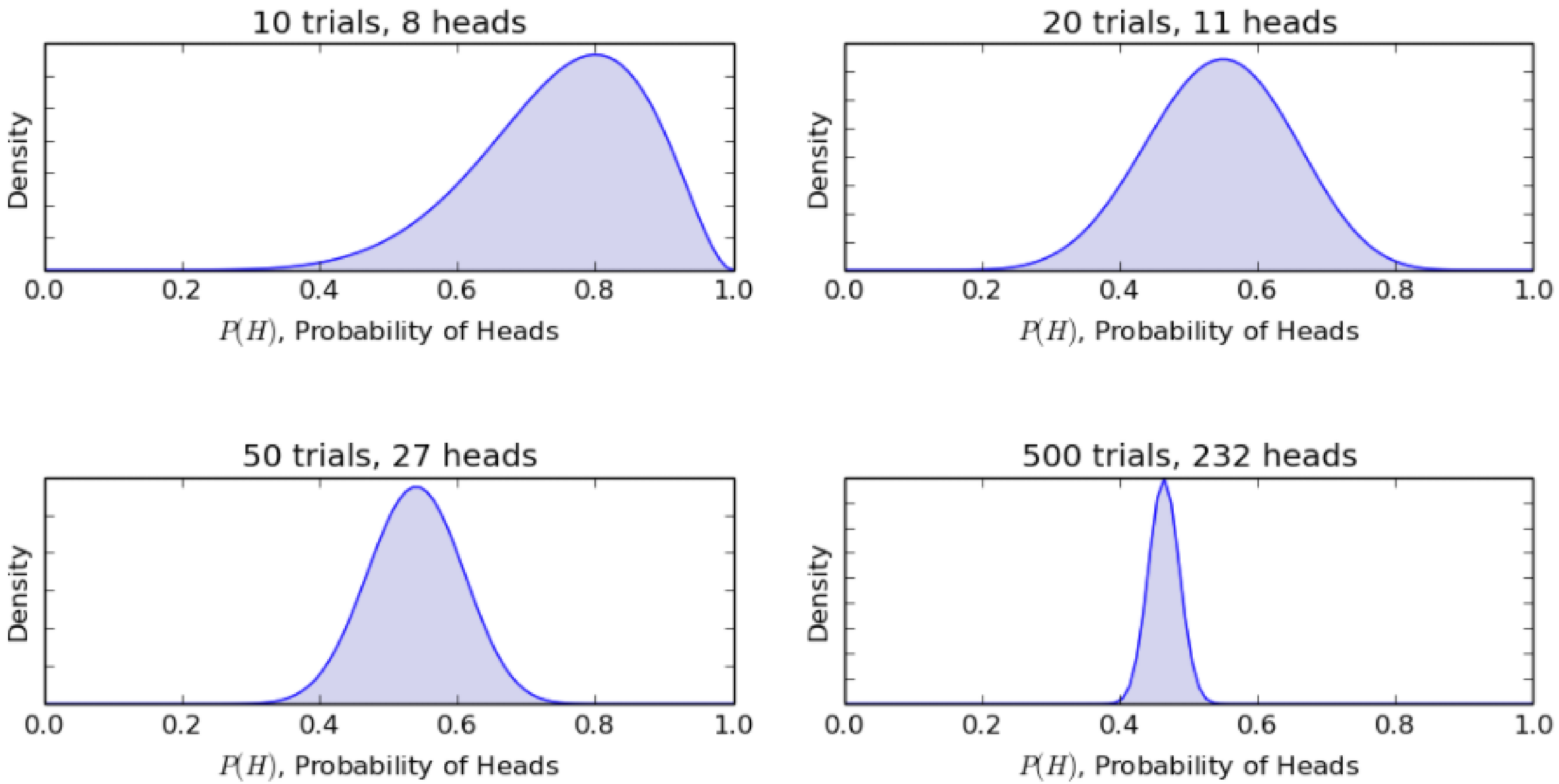
Likelihood $\boldsymbol{p(\mathcal{D}|\theta)}$ → Likelihood of observing $\boldsymbol{\mathcal{D}}$ given $\boldsymbol{\theta}$

Posterior $\boldsymbol{p(\theta|\mathcal{D})}$ → Believe of parameters after observing data

?

Example

MLTS Exercise 02



Thank you for your attention!

Any questions?

Practice Questions

MLTS Exercise 02

- **Can the following function be considered in a linear regression:**

$$y = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$$

  - Yes
  - No

- **Why might we want to employ a Bayesian instead of a Frequentists model in a safety-critical environment?**
  - The Bayesian model gives us more accurate predictions
  - The Bayesian model is less biased
  - The Bayesian model gives us an estimate for the uncertainty of our predictions

- **What is a prior probability in Bayes theorem?**
  - Our estimate after considering the data and the likelihood
  - Our assumptions about parameters e.g. the range of values a parameter can take
  - Prior knowledge we can include into the model
- **What is a reasonable prior theta in the coin flip example?**
  - The results will be the same after a finite number of trials and independent of the starting theta, as we will update our believe after observing data
  - Theta = 0.75
  - Theta = 0.5
  - Theta = 0.25