Linear Regression
Practicals Machine Learning 1, SS22

Ceca Kraišniković
Institute of Theoretical Computer Science
Graz University of Technology, Austria

A few things to be aware of…

- A course with many participants:
  - Educational background / previous knowledge might be different
  - Your expectations
  - Previous experiences

A few things to be aware of…

- A course with many participants:
  - Educational background / previous knowledge might be different
  - Your expectations
  - Previous experiences

- **Diversity** (@TUG)**:** gender, ethnic origin, age, cultural background, sexual orientation, language, skills, knowledge, mobility needs, lifestyles…

A few things to be aware of…

- A course with many participants:
  - Educational background / previous knowledge might be different
  - Your expectations
  - Previous experiences

- **Diversity** (@TUG)**:** gender, ethnic origin, age, cultural background, sexual orientation, language, skills, knowledge, mobility needs, lifestyles…

- Expected from ALL of us:
  - Fair communication & cooperation
  - Respectful participation (especially among fellow colleagues)
  - Avoiding any form of discrimination!

A few things to be aware of…

- A course with many participants:
  - Educational background / previous knowledge might be different
  - Your expectations
  - Previous experiences

- **Diversity** (@TUG)**:** gender, ethnic origin, age, cultural background, sexual orientation, language, skills, knowledge, mobility needs, lifestyles…

- Expected from ALL of us:
  - Fair communication & cooperation
  - Respectful participation (especially among fellow colleagues)
  - Avoiding any form of discrimination!

- Prefer communication in person (instead of writing emails for solving “big problems”)
- Content:
  - Often a recap of the lecture + examples in Python (or small exercises to do by hand)
  - Goal: a different perspective, but often not complete $\Rightarrow$ Attend the lectures!

Recap: Matrix inverse & Pseudoinverse (1)

- **Inverse matrix** (of matrix A) is denoted as A^{-1}, defined only for square matrices, and with the property

$$AA^{-1} = A^{-1}A = I,$$

where I is the identity matrix. If A^{-1} exists, it is unique.

- For solving a system of linear equations

$$\begin{aligned} Ax &= b \\ A^{-1}Ax &= A^{-1}b \\ Ix &= A^{-1}b \\ x &= A^{-1}b \end{aligned}$$

- Pseudoinverse matrix (the best fit in the least-squares sense, for overdetermined system):

$$\begin{aligned} Ax &= b \\ A^TAx &= A^Tb \\ x &= (A^TA)^{-1}A^Tb \end{aligned}$$

Pseudoinverse: $A^+ = (A^TA)^{-1}A^T$

Recap: Matrix inverse & Pseudoinverse (2)

- Matrix inverse

$$A^{-1} = \frac{1}{|A|}(\mathrm{adj}(A)),$$

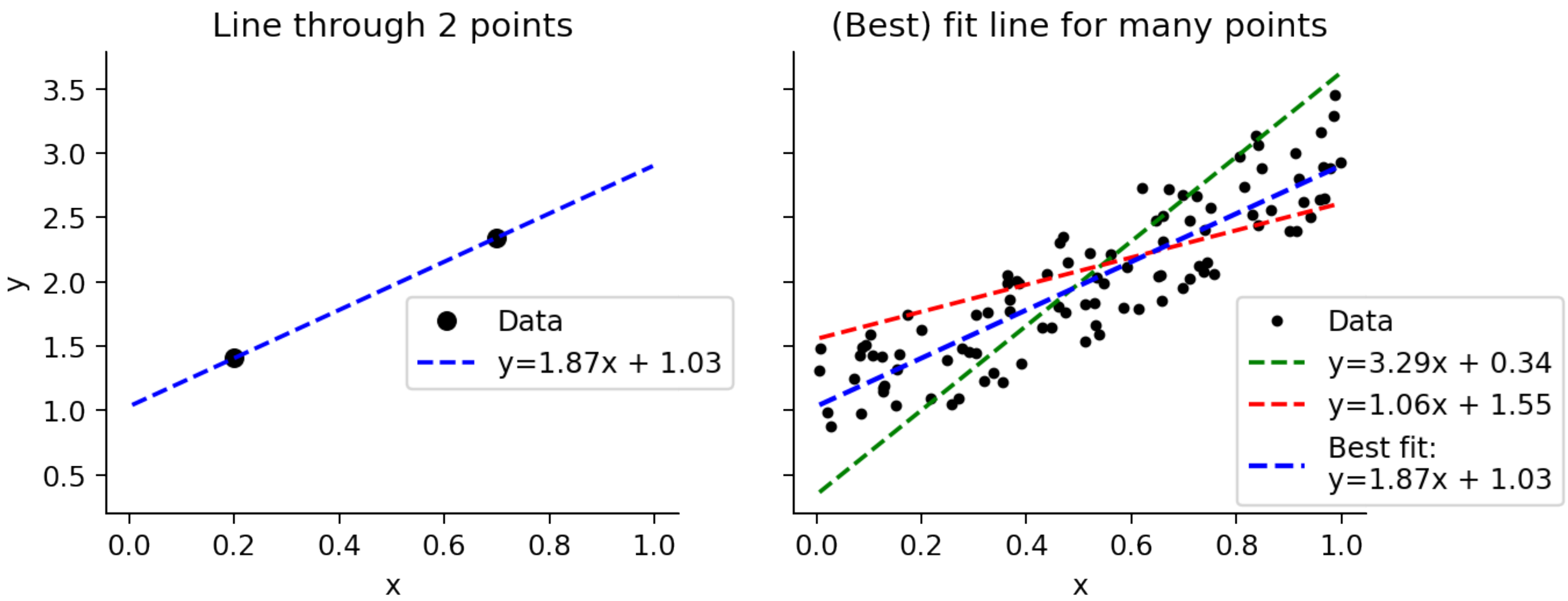
with adj(A) being the transpose of its cofactor matrix.

- A^{-1} of a 2-by-2 matrix easily calculated:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad A^{-1} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

where $|A|$ is determinant of A.

Introduction to Linear Regression (1)



Introduction to Linear Regression (2)

- A linear regression model:
  - fits the existing data in the best possible way,
  - predicts values when independent variables are given.

Introduction to Linear Regression (2)

- A linear regression model:
  - fits the existing data in the best possible way,
  - predicts values when independent variables are given.

- To estimate the parameters of the model, the following components are necessary:
  - a data set consisting of m data points, e.g., $\{(x_1, y_1), \ldots, (x_m, y_m)\}$
  - an assumption about the appropriate "shape" of the function to use for fitting the model to the data (also known as "hypothesis" or "basis function"),
  - a criteria to measure the best fit.

Introduction to Linear Regression (3): Choosing an appropriate hypothesis

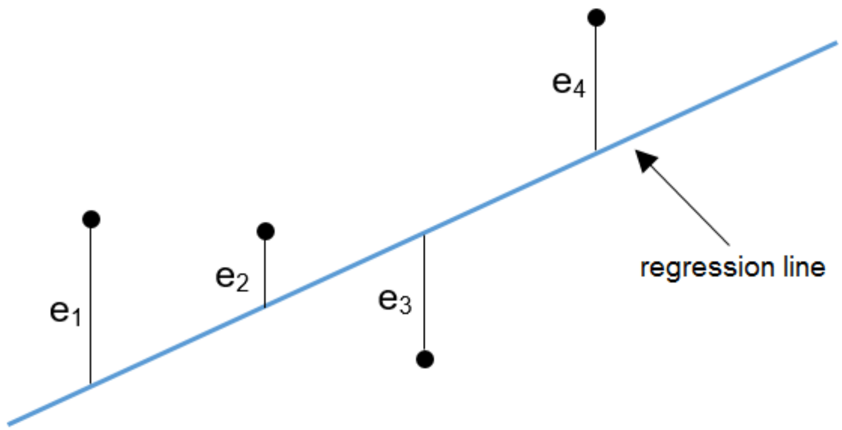
- The choice depends completely on the modeler
- The common choices are:
  - Linear hypothesis $h_\theta(x) = \theta_0 + \theta_1 x$
  - Polynomial hypothesis $h_\theta(x) = \sum_{k=0}^{n} \theta_k x^k$
  - Non-linear hypothesis $h_\theta(x) = \sum_{k=0}^{n} \theta_k g_k(x)$, where $g_k(x)$ are chosen beforehand, e.g., $\sin(x)$, $\cos(x)$, Gaussian basis function $\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

Introduction to Linear Regression (3): Choosing an appropriate hypothesis

- The choice depends completely on the modeler
- The common choices are:
  - Linear hypothesis $h_\theta(x) = \theta_0 + \theta_1 x$
  - Polynomial hypothesis $h_\theta(x) = \sum_{k=0}^{n} \theta_k x^k$
  - Non-linear hypothesis $h_\theta(x) = \sum_{k=0}^{n} \theta_k g_k(x)$, where $g_k(x)$ are chosen beforehand,

    e.g., $\sin(x)$, $\cos(x)$, Gaussian basis function $\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.
- For the assumed hypothesis, the parameters θ_k are determined through the minimization of an error function on the data set.
- **Question:** Why is it called linear, when we also use non-linear transformations of features x?

Introduction to Linear Regression (4): Measuring the best fit



- *Error* function, or *cost* function appropriate to use - Mean Squared Error (MSE):

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 = \frac{1}{m} \sum_{i=1}^{m} e_i^2$$

- To find $\theta = [\theta_0, \theta_1, \ldots, \theta_n]^T$, the error function can be minimized using *Gradient Descent*, or the solution can be found analytically — a closed-form solution for linear regression problems exists.

Exercise (to be done after the second part of the tutorial, *Linear_regression_recap.pdf*)

Consider a linear regression problem. For simplicity, given are only two points, and we want to fit a line through them. The given points are: $(x_1, y_1) = (1,4)$, and $(x_2, y_2) = (2,5)$. We want to find $y = f(x)$.

First, specify vectors x and y for these two points.

Form an appropriate design matrix X that contains zero feature, and the feature x. (The dimension of this matrix must be 2-by-2. **Why?**)

Write down the expression that represents the analytical solution for linear regression problem to find optimal parameters θ^*, and calculate θ^* using that expression.

Finally, sketch the two points in a coordinate system, and draw a line with the coefficients that you calculated (parameters θ^*).
What is the slope of this line? What is the intercept?