Lecture 3a. Linear Regression – Decision Theory

COMP90051 Statistical Machine Learning

Semester 2, 2020
Lecturer: Ben Rubinstein

This lecture

- **Linear regression**
  - Simple model (convenient maths at expense of flexibility)
  - Often needs less data, “interpretable”, lifts to non-linear
  - Derivable under all Statistical Schools: Lect 2 case study
    - This week: Frequentist + **Decision theory derivations**
  - Later in semester: Bayesian approach
  - Convenient optimisation: Training by “analytic” (exact) solution

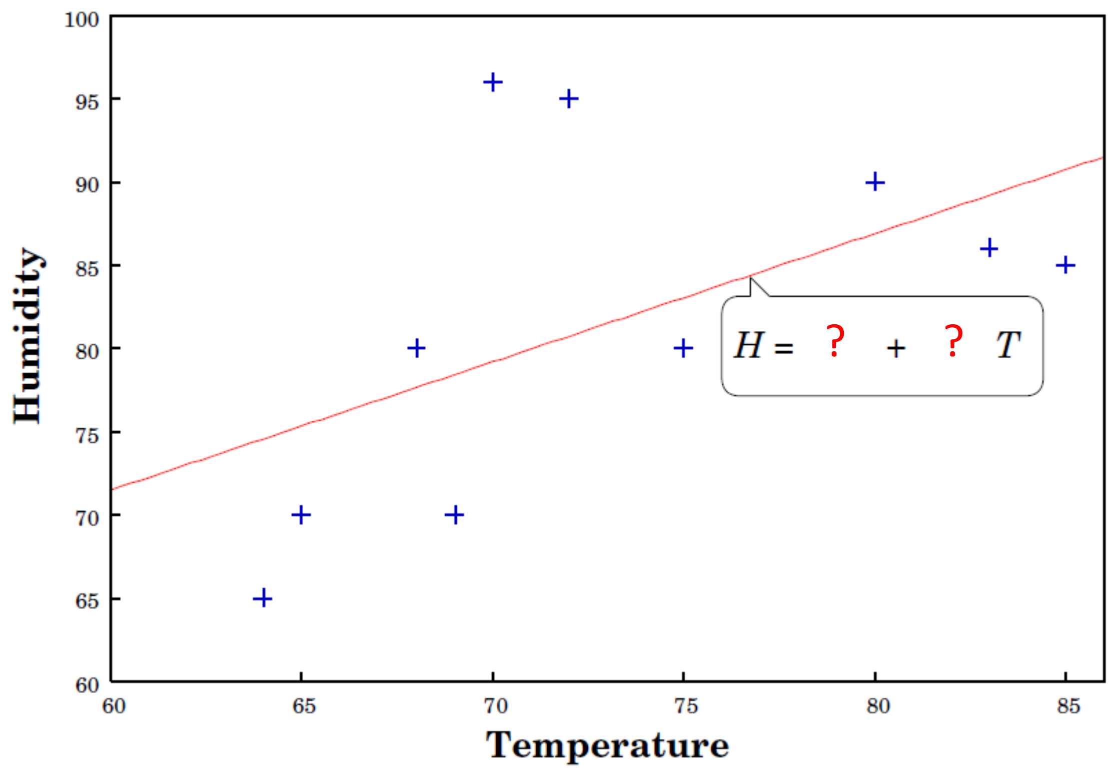
- Basis expansion: Data transform for more expressive models

Linear Regression via Decision Theory

A warm-up example

Example: Predict humidity from temperature

Temperature	Humidity
TRAINING DATA	
85	85
80	90
83	86
70	96
68	80
65	70
64	65
72	95
69	70
75	80
TEST DATA	
75	70



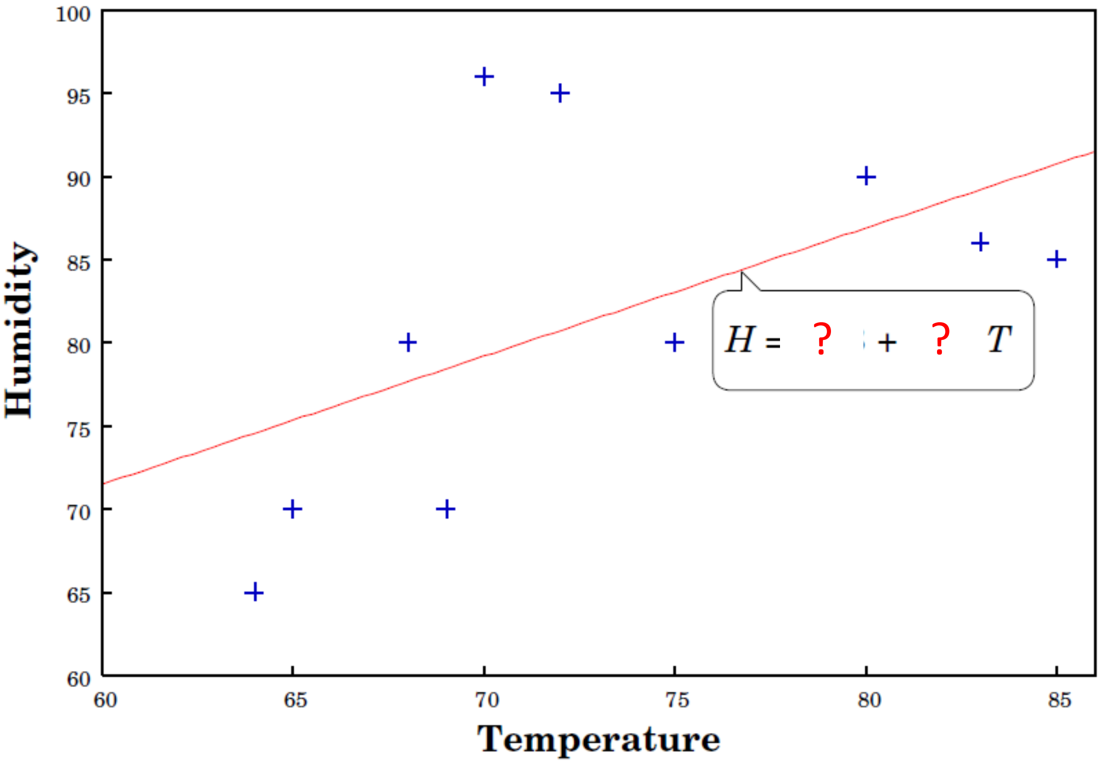
In regression, the task is to predict numeric response (*aka* dependent variable) from features (*aka* predictors or independent variables)

Assume a linear relation: $H = a + bT$

(H – humidity; T – temperature; a, b – parameters)

Example: Problem statement

- The model is
  $$H = a + bT$$
- Fitting the model = finding "best" a, b values for data at hand
- Important criterion: minimise the sum of squared errors (*aka* residual sum of squares)



Example: Minimise Sum Squared Errors

To find a, b that minimise $L = \sum_{i=1}^{10}\left(H_i - (a + b\,T_i)\right)^2$

set derivatives to zero:

$$\frac{\partial L}{\partial a} = -2\sum_{i=1}^{10}(H_i - a - b\,T_i) = 0$$

if we know b, then $\hat{a} = \frac{1}{10}\sum_{i=1}^{10}(H_i - b\,T_i)$

$$\frac{\partial L}{\partial b} = -2\sum_{i=1}^{10}T_i(H_i - a - b\,T_i) = 0$$

if we know a, then $\hat{b} = \frac{1}{\sum_{i=1}^{10} T_i^2}\sum_{i=1}^{10} T_i(H_i - a)$

High-school optimisation:
- Write derivative
- Set to zero
- Solve for model
- (Check 2nd derivatives)

Can we be more systematic?

Example: Analytic solution

- We have two equations and two unknowns a, b
- Rewrite as a system of linear equations

$$\begin{pmatrix} 10 & \sum_{i=1}^{10} T_i \\ \sum_{i=1}^{10} T_i & \sum_{i=1}^{10} {T_i}^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{10} H_i \\ \sum_{i=1}^{10} T_i H_i \end{pmatrix}$$

- Analytic solution: $a = 25.3, b = 0.77$
- (Solve using `numpy.linalg.solve` or sim.)

More general decision rule

- Adopt a linear relationship between response $y \in \mathbb{R}$ and an instance with features $x_1, \dots, x_m \in \mathbb{R}$

$$\hat{y} = w_0 + \sum_{i=1}^{m} x_i w_i$$

Here $w_0, \dots, w_m \in \mathbb{R}$ denote weights (model parameters)

- Trick: add a dummy feature $x_0 = 1$ and use vector notation

$$\hat{y} = \sum_{i=0}^{m} x_i w_i = \boldsymbol{x}'\boldsymbol{w}$$

A lowercase symbol in **bold face** indicates a vector; $\boldsymbol{x}'$ denotes transpose

Summary

- Linear regression
  - Simple, effective, “interpretable”, basis for many approaches
  - Decision-theoretic frequentist derivation

Next time:

Frequentist derivation; Solution/training approach