Regresi Linier

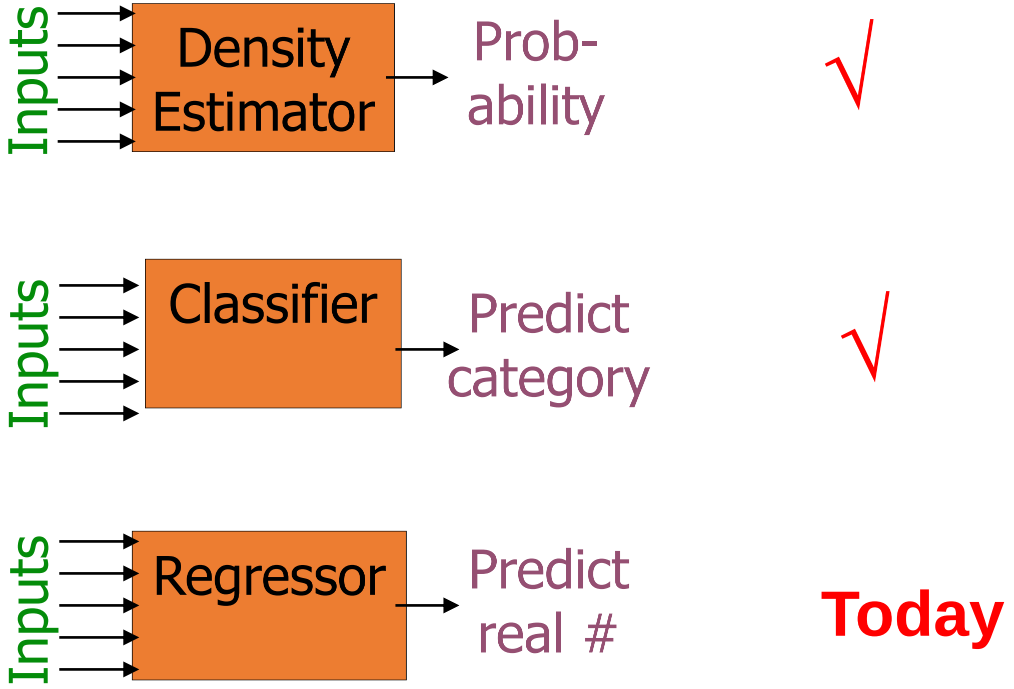
Outline

- Regression *vs* Classification
- Linear regression – another discriminative learning method
  - As optimization ➔ Gradient descent
  - As matrix inversion (<u>O</u>rdinary <u>L</u>east <u>S</u>quares)
- Overfitting and bias-variance

- Bias-variance decomposition for classification

What is regression?

Where we are


Inputs
Density Estimator
Prob-ability
√
Inputs
Classifier
Predict category
√
Inputs
Regressor
Predict real #
Today

Regression examples

Stock market

Weather prediction

Temperature
72° F

Predict the temperature at any given location

Prediction of menu prices

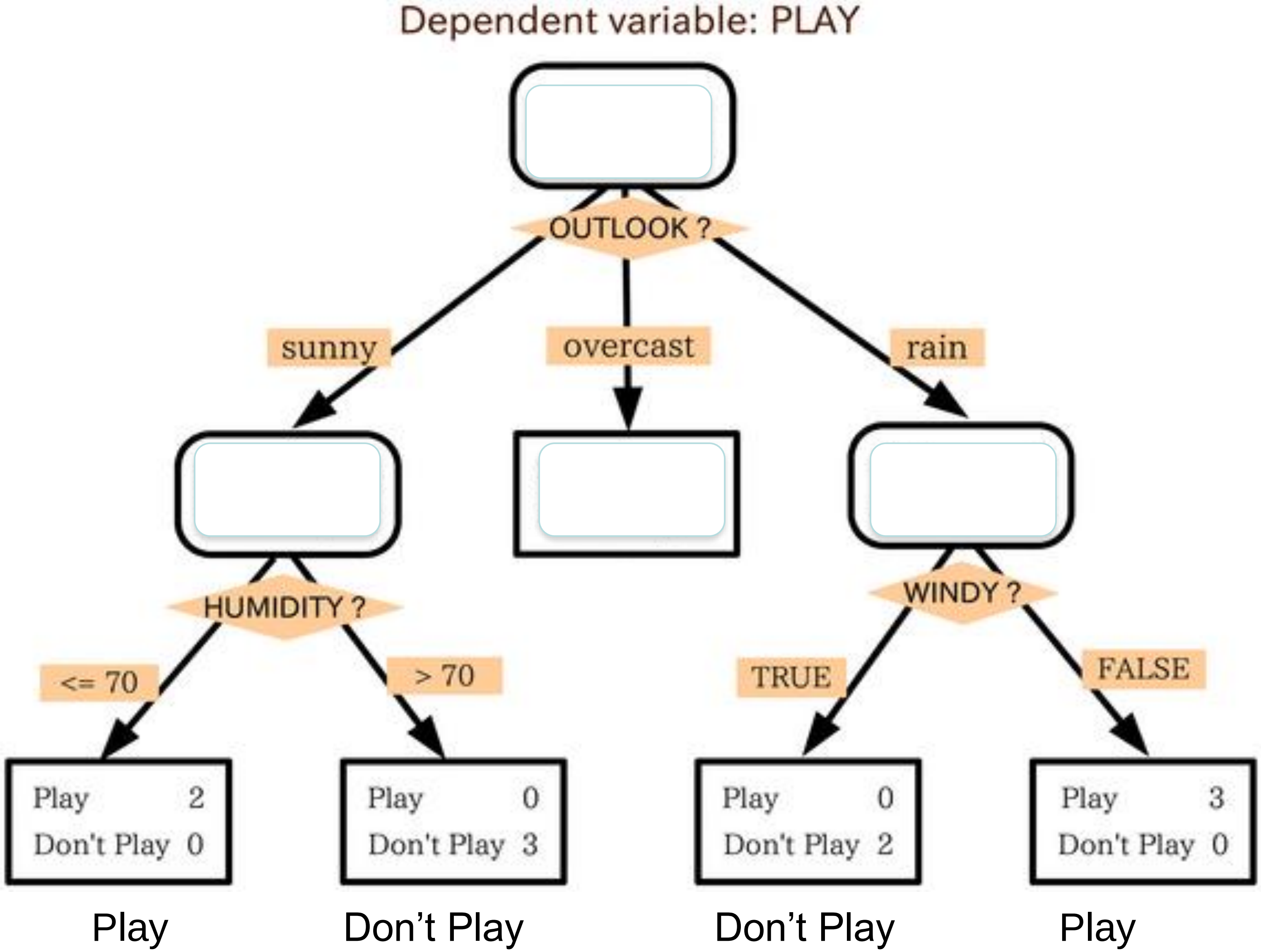
Chaheau Gimpel … and Smith EMNLP 2012

(a) METADATA: ambience	
dive-y	-0.015
intimate	-0.013
trendy	-0.012
casual	-0.005
romantic	-0.004
classy	-7e-6
touristy	0.058
upscale	0.099

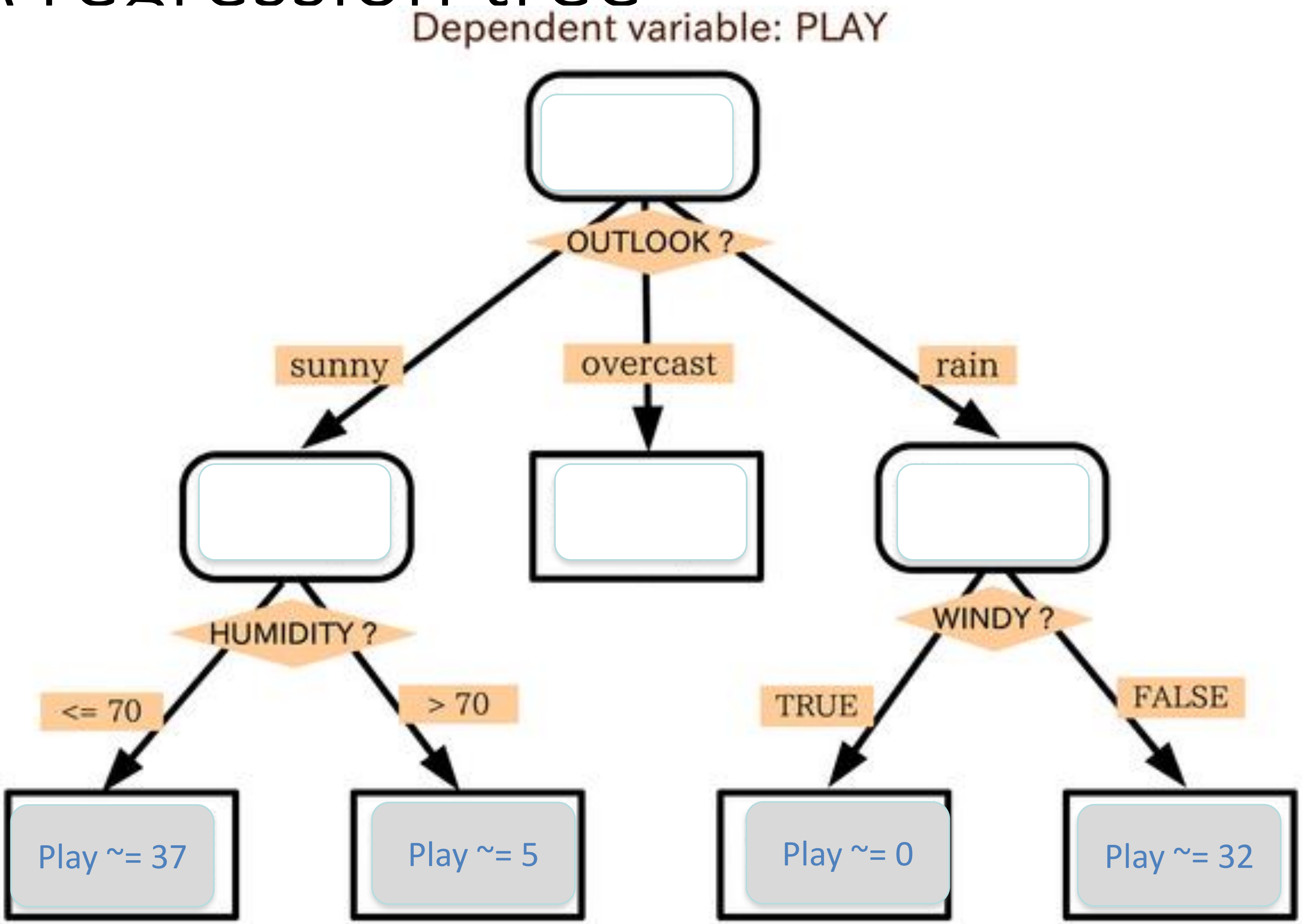
(d) MENUDESC: _ = "of chicken"	
slices _	-0.102
bits _	-0.032
cubes _	-0.030
pieces _	-0.024
strips _	-0.001
chunks _	0.015
morsels _	0.025
pcs _	0.040
cuts _	0.042

(c) MENUDESC: descriptors	
old time favorite	-0.112
fashioned	-0.034
…	
artisanal	0.064
raised	0.066
heirloom	0.083
wild	0.084
hormone	0.085
farmed	0.099
hand picked	0.101
wild caught	0.116
farmhouse	0.133

A decision tree: classification



A regression tree



Play = 30m, 45min Play = 0m, 0m, 15m Play = 0m, 0m Play = 20m, 30m, 45m,

Theme for the week: learning as optimization

Types of learners

- Two types of learners:

  1. Generative: make assumptions about how to generate data (given the class)

     - e.g., naïve Bayes

  2. Discriminative

     - directly estimate a decision rule/boundary
     - e.g., logistic regression

Today: another discriminative learner, but for regression tasks

Least Mean Squares

Regression for LMS as optimization

Toy problem #2

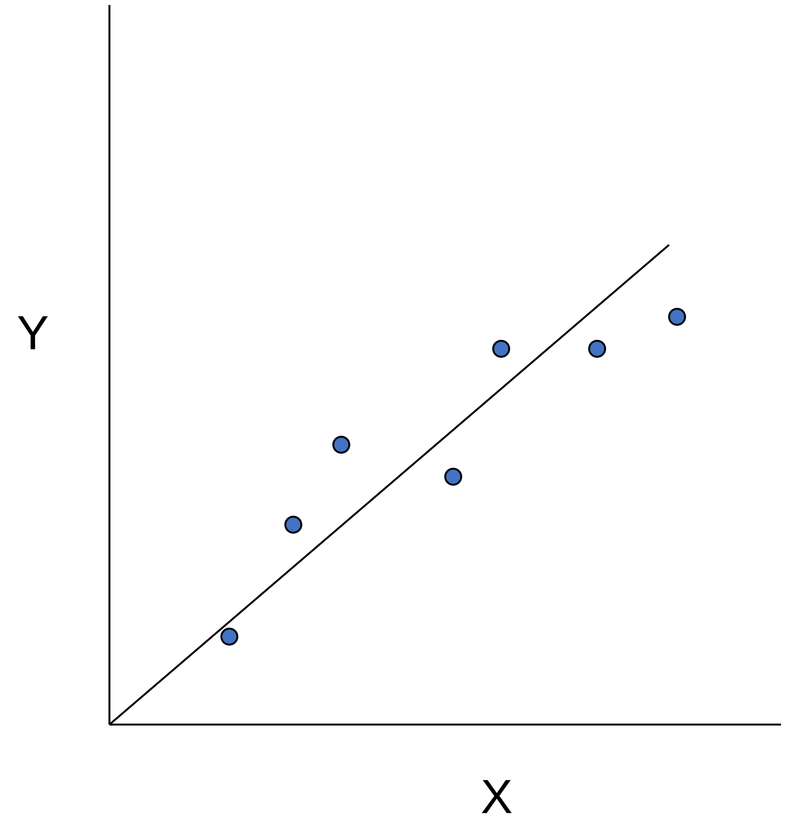
Linear regression

- Given an input x we would like to compute an output y
- For example:

  - Predict height from age

  - Predict Google's price from Yahoo's price

  - Predict distance from wall from sensors



Linear regression

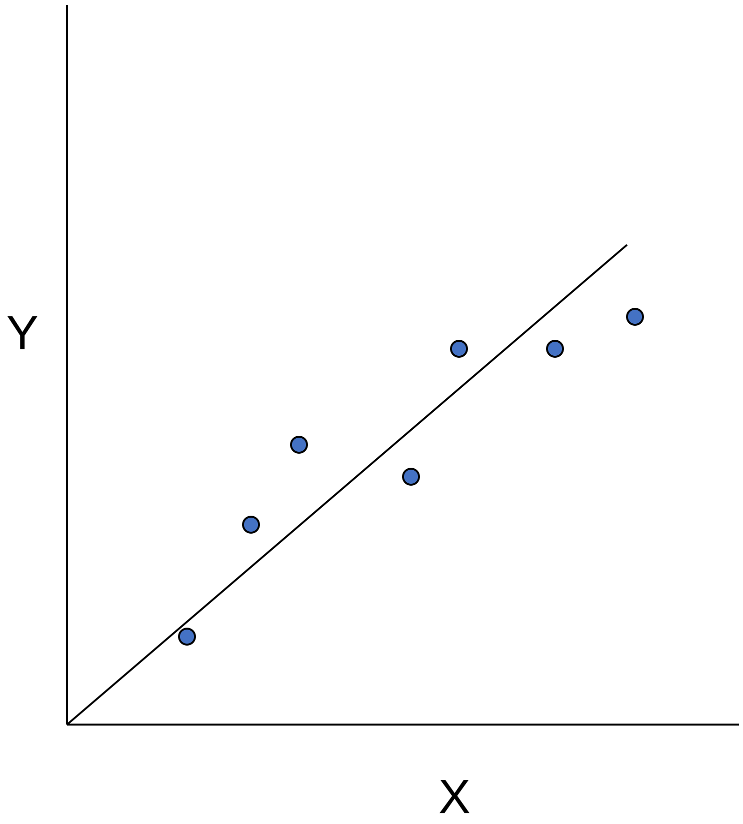
- Given an input x we would like to compute an output y
- In linear regression we assume that y and x are related with the following equation:

What we are trying to predict

Observed values

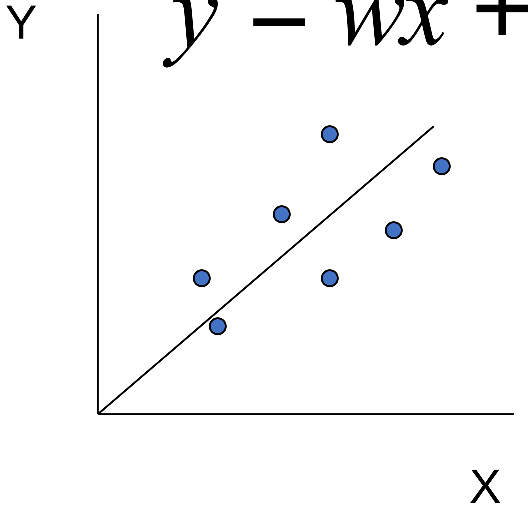
$$y = wx+\varepsilon$$

where w is a parameter and ε represents measurement or other noise



Linear regression

$$y = wx + e$$



- Our goal is to estimate *w* from a training data of $\langle x_i, y_i \rangle$ pairs
- Optimization goal: minimize squared error (least squares):

$$\arg\min_w \sum_i (y_i - wx_i)^2$$

- Why least squares?

  - minimizes squared distance between measurements and predicted line

  - has a nice probabilistic interpretation

  - the math is pretty

see HW

Solving linear regression

- To optimize:
- We just take the derivative w.r.t. to w ….

prediction

$$\frac{\partial}{\partial w} \sum_i (y_i - w x_i)^2 = 2 \sum_i -x_i (y_i - w x_i)$$

prediction

Compare to logistic regression…

$$\frac{\partial}{\partial w^j} \log P(Y = y | X = \mathbf{x}, \mathbf{w}) = (y - p) x^j$$

Solving linear regression

- To optimize – closed form:
- We just take the derivative w.r.t. to w and set to 0:

$$\frac{\partial}{\partial w}\sum_i (y_i - wx_i)^2 = 2\sum_i -x_i(y_i - wx_i) \Rightarrow$$

$$2\sum_i x_i(y_i - wx_i) = 0 \quad \Rightarrow \quad 2\sum_i x_i y_i - 2\sum_i w x_i x_i = 0$$

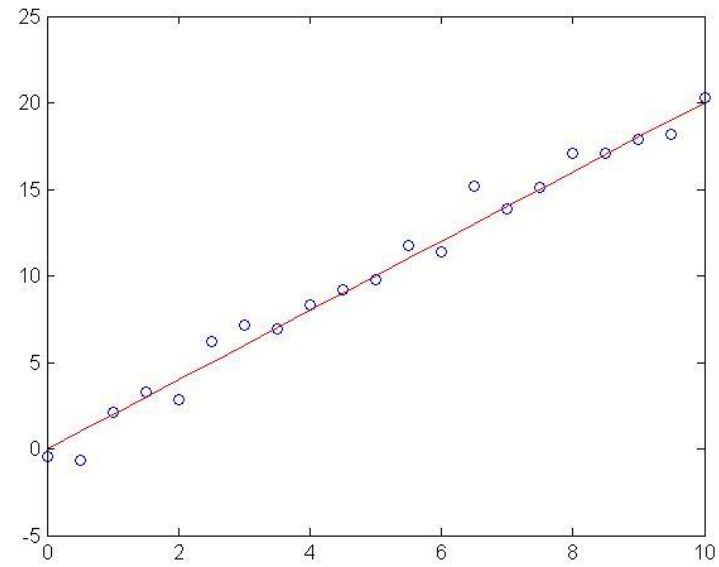
$$\sum_i x_i y_i = \sum_i w x_i^2 \quad \Rightarrow$$

$$w = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

covar(X,Y)/var(X)
if mean(X)=mean(Y)=0

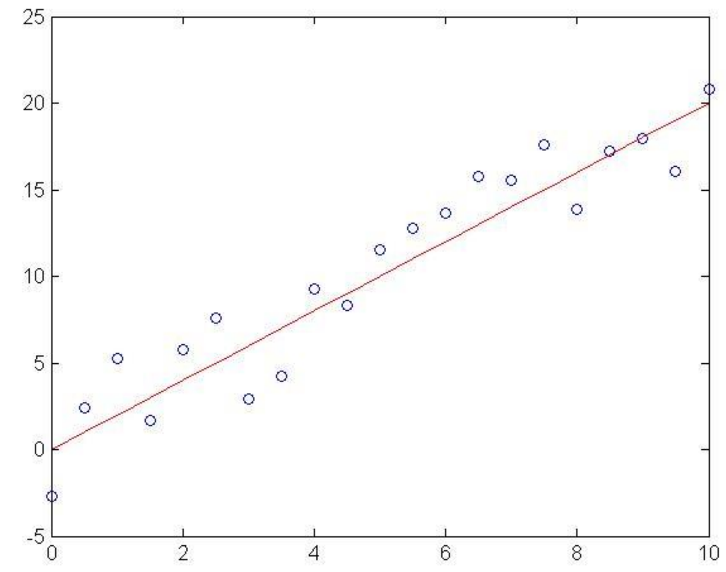
Regression example

- Generated: w=2
- Recovered: w=2.03
- Noise: std=1



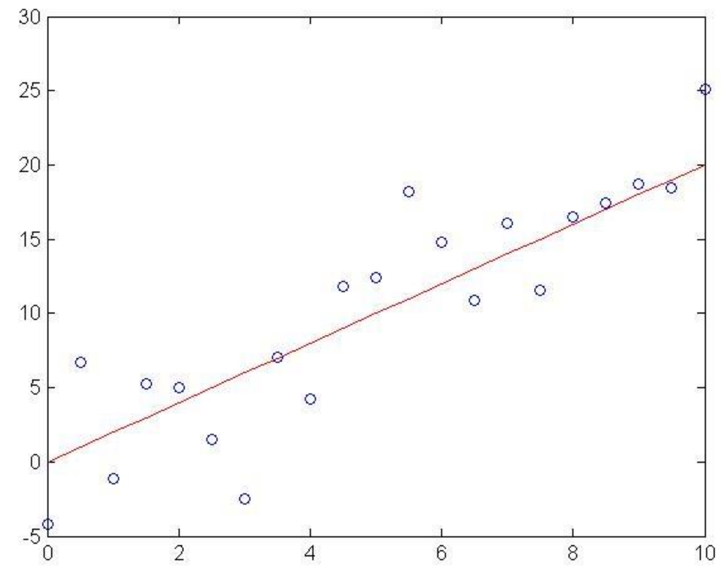
Regression example

- Generated: w=2
- Recovered: w=2.05
- Noise: std=2



Regression example

- Generated: w=2
- Recovered: w=2.08
- Noise: std=4

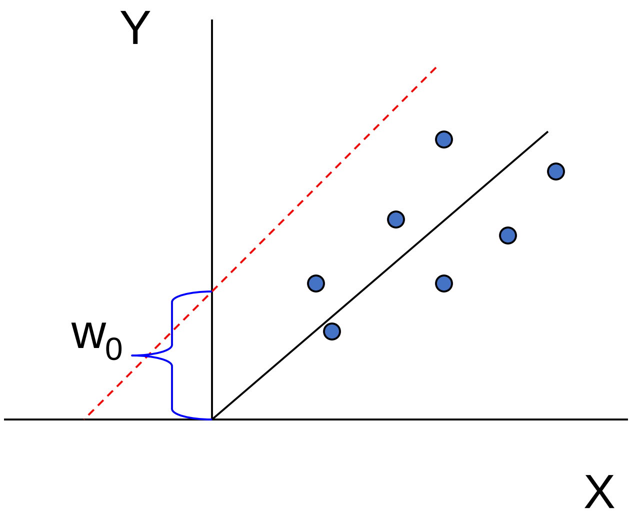


Bias term

- So far we assumed that the line passes through the origin
- What if the line does not?
- No problem, simply change the model to

$$y = w_0 + w_1 x + \varepsilon$$

- Can use least squares to determine w_0 , w_1



$$w_0 = \frac{\sum_i y_i - w_1 x_i}{n}$$

$$w_1 = \frac{\sum_i x_i (y_i - w_0)}{\sum_i x_i^2}$$

Multivariate regression

- What if we have several inputs?

  - Stock prices for Yahoo, Microsoft and Ebay for the Google prediction task

- This becomes a multivariate regression problem
- Again, its easy to model:

$$y = w_0 + w_1 x_1 + \ldots + w_k x_k + \varepsilon$$

Google's stock price

Yahoo's stock price

Microsoft's stock price

Multivariate regression

- What if we have several inputs?

  - Stock prices for Yahoo, Microsoft and Ebay for the Google prediction task
- This becomes a multivariate regression problem
- Again, its easy to model:

$$y = w_0 + w_1x_1 + \ldots + w_kx_k + \varepsilon$$

Not all functions can be approximated by a line/hyperplane…

$$y=10+3x_1^2-2x_2^2+\varepsilon$$

In some cases we would like to use polynomial or other terms based on the input data, are these still linear regression problems?

Yes. As long as the *coefficients* are linear the equation is still a linear regression problem!

Non-Linear basis function

- So far we only used the observed values $x_1, x_2, \ldots$
- However, linear regression can be applied in the same way to **functions** of these values
  - Eg: to add a term w $x_1 x_2$ add a new variable $z = x_1 x_2$ so each example becomes: x_1, x_2, …. z
- As long as these functions can be directly computed from the observed values the parameters are still linear in the data and the problem remains a multi-variate linear regression problem

$$y = w_0 + w_1 x_1^2 + \ldots + w_k x_k^2 + \varepsilon$$

Non-Linear basis function

- How can we use this to add an intercept term?

Add a new "variable" $z=1$ and weight w_0

Non-linear basis functions

- What type of functions can we use?
- A few common examples:

- Polynomial: $\phi_j(x) = x^j$ for j=0 … n

- Gaussian: $$f_j(x) = \frac{(x - m_j)}{2s_j^2}$$

- Sigmoid: $$f_j(x) = \frac{1}{1 + \exp(- s_j x)}$$

- Logs: $$f_j(x) = \log(x+1)$$

Any function of the input values can be used. The solution for the parameters of the regression remains the same.

General linear regression problem

- Using our new notations for the basis function linear regression can be written as

$$y = \sum_{j=0}^{n} w_j \phi_j(x)$$

- Where $\phi_j(\mathbf{x})$ can be either x_j for multivariate regression or one of the non-linear basis functions we defined
- … and $\phi_0(\mathbf{x})=1$ for the intercept term

Contoh Soal

- Perhatikan data Biaya iklan yang digunakan (X) dan hubungannya dengan Tingkat penjualan (Y) diberikan dalam dataset berikut :

No	X	Y
1	41	1250
2	54	1380
3	63	1425
4	54	1425
5	48	1450
6	46	1300
7	62	1400
8	61	1510
9	64	1575
10	71	1650

Tentukan persamaan Regresinya!

- Penyelesaian :
  - Mengestimasi least squares/kuadrat terkecil dari koefisien Regresi :

$$n = 10 \qquad \sum x = 564 \qquad \sum x^2 = 32604$$

$$\sum y = 14365 \qquad \sum xy = 818755$$

$$b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{10(818755) - (564)(14365)}{10(32604) - (564)^2} = 10.8$$

$$b_0 = \frac{\left(\sum y - b_1 \sum x\right)}{n} = 1436.5 - 10.8(56.4) = 828$$

$$\hat{y} = b_0 + b_1 x$$

  - Hasil Estimasi Persamaan Regresinya adalah :

$$\hat{y} = 828 + 10.8x$$

Ini berarti bahwa jika biaya iklan meningkat sebesar \$ 1, maka kita akan mendapati tingkat penjualan naik \$ 10.8

Latihan

- Tabel menyajikan data dengan variabel X adalah umur mobil dan variabel Y adalah harga.

Age (yrs) x	Price ($100s) y
5	85
4	103
6	70
5	82
5	89
5	98
6	66
6	95
2	169
7	70
7	48