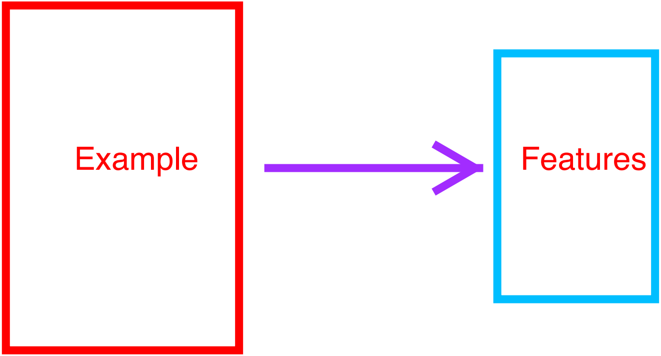
Machine Learning basics

E : experience

T : task

P : performance


Example
Features

Classification

input

⋮

$\downarrow$

features

$\vec{x}$

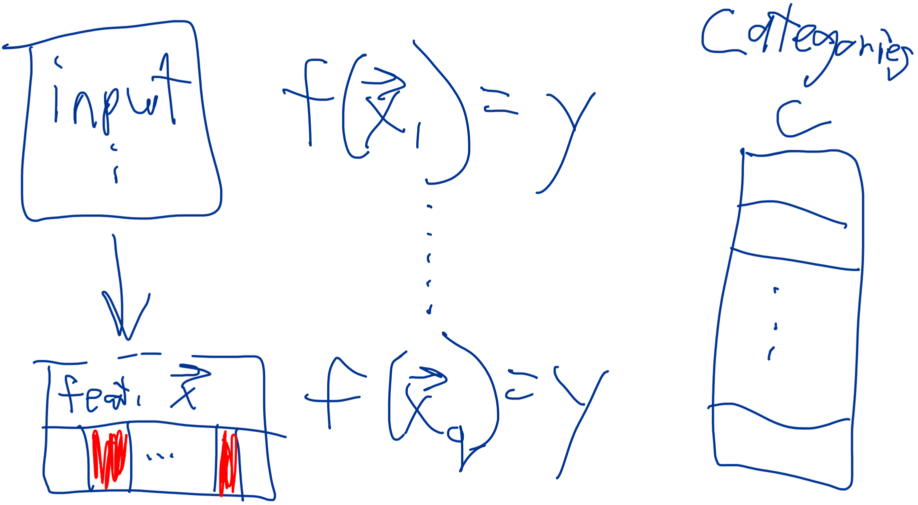
$$f(\vec{x}) = y$$

Categories

C

⋮

Classification w. missing input



$$f(\vec{x}_1) = y$$

$$\vdots$$

$$f(\vec{x}_q) = y$$

Regression

$$\boxed{\text{input}}$$

$$\downarrow$$

$$\boxed{\text{features } \vec{x}}$$

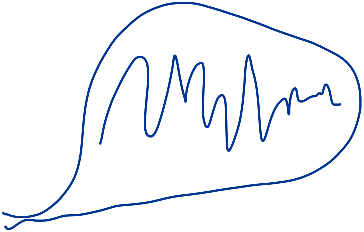
$$f(\vec{x}) = y$$

Range: $\mathbb{R}$

OCR

61801 $\Rightarrow$ 61801

ASR



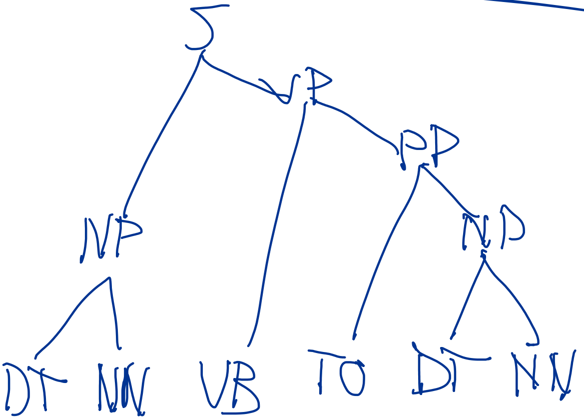
⇒ iɣaquɬχət

MT

Akuzilleput Igaqullghet

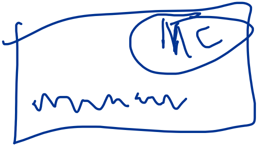
Our words put to paper

Structure Prediction



The boy walks to the park

Anomaly detection

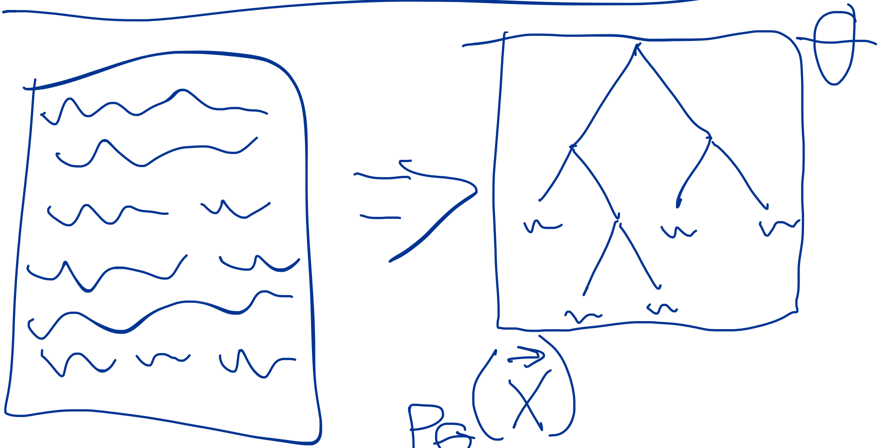


Supervised vs Unsupervised Learning

Unsupervised

Estimate $p(\vec{x})$ from unlabelled examples

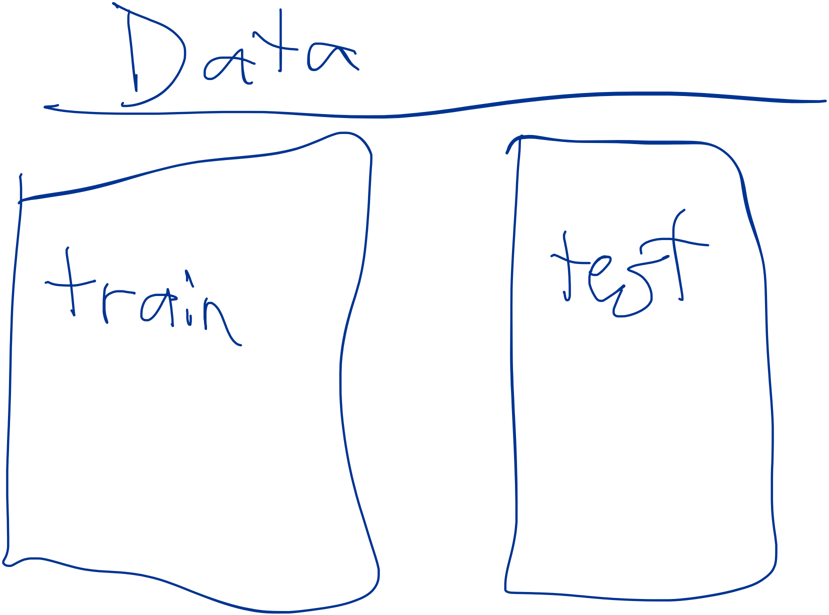
Unsupervised grammar induction



$p_\theta(\vec{x})$

Supervised grammar induction

$p_\theta(y \mid x)$


Data
train
test

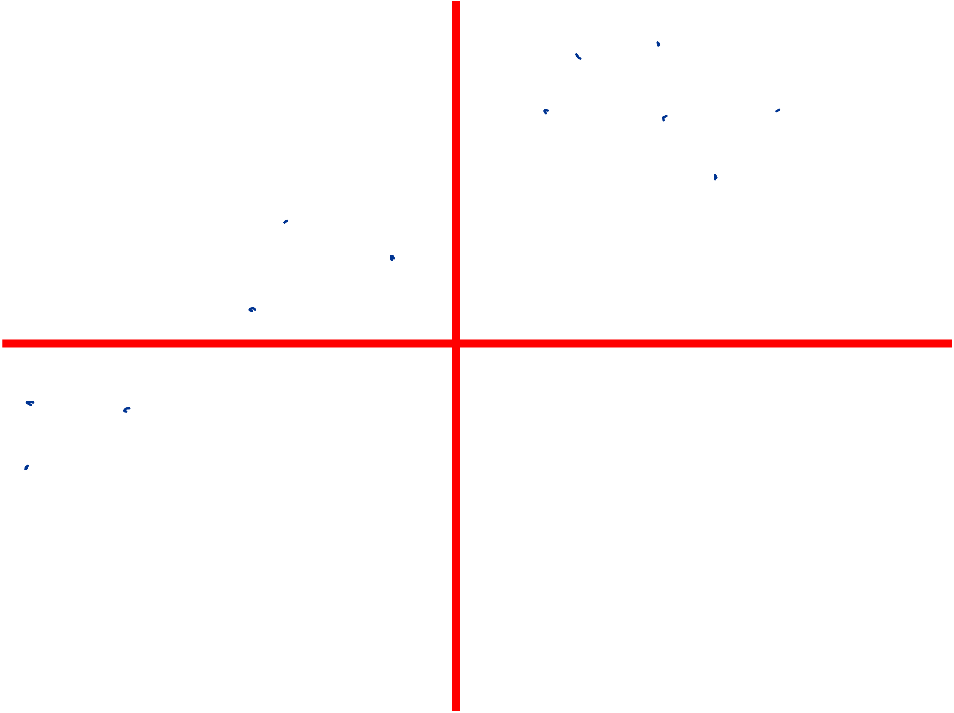
Goal:

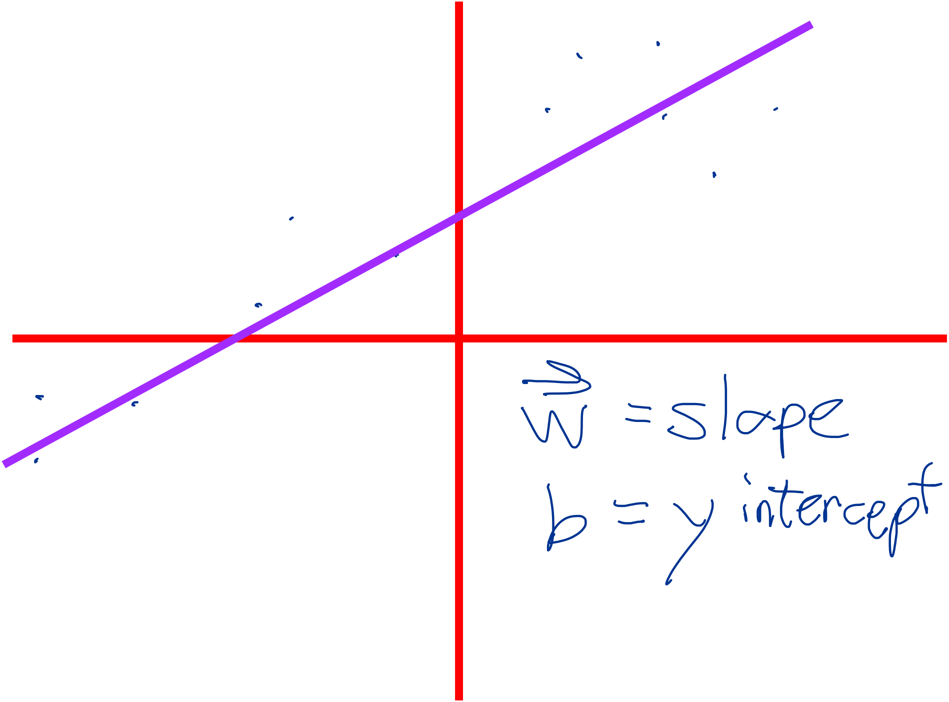
Learn a model over training data that generalizes to unseen test data

Linear regression

$$\hat{y} = \vec{w}\vec{x} + b$$

Learn $\vec{w}$ ($\cancel{b}$ b)




$\vec{w}$ = slope
b = y intercept

Linear regression (as shown in the text) can be solved in closed form

However, most problems we are interested in cannot be solved in closed form

Underfitting

&

Overfitting

(see Fig 5.2)

Maximum Likelihood

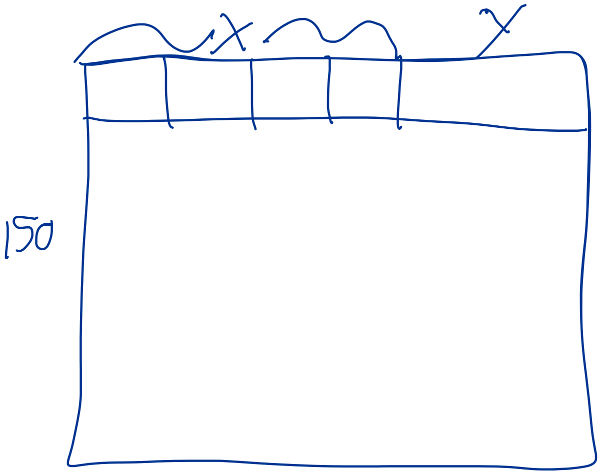
$$\hat{\theta} = \arg\max_{\theta} P(X, \theta)$$

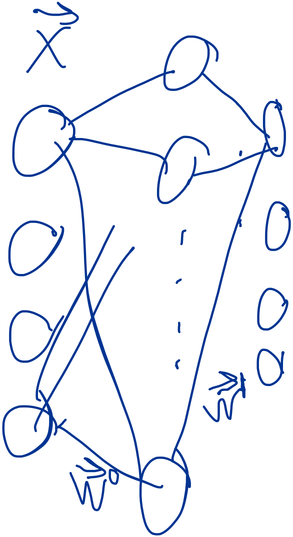
where X is set of training examples

Example:

Train an n-gram language model from training data

The boy walks to the park





hidden layer

nodes

σ \eta η