

Strojové učenie

Ľahké intro



Ondrej Kaššák

28th September 2017

Motivácia

Čo je strojové učenie?

3

- Ľudia - Správanie sa na základe naučených skúsenosti
- Počítače - Nasledovanie inštrukcií

Čo je strojové učenie?

4

- Ľudia - Správanie sa na základe **naučených skúsenosti**
- Počítače - Nasledovanie **inštrukcií**

Ako to vieme spojiť?

- Počítače - Správanie sa na základe **dát**

- Inštancie (pozorovania)

- Atribúty

- Numerické
- Ordinálne
- Nominálne (kategorické)

- Vstupné (X)
- Výstupné (Y)

- Dáta

- Trénovacie
- Testovacie

	Sepal.L	Sepal.W	Petal.L	Petal.W	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
51	7.0	3.2	4.7	1.4	versicolor
52	6.4	3.2	4.5	1.5	versicolor
53	6.9	3.1	4.9	1.5	versicolor
54	5.5	2.3	4.0	1.3	versicolor
55	6.5	2.8	4.6	1.5	versicolor
101	6.3	3.3	6.0	2.5	virginica
102	5.8	2.7	5.1	1.9	virginica
103	7.1	3.0	5.9	2.1	virginica
104	6.3	2.9	5.6	1.8	virginica
105	6.5	3.0	5.8	2.2	virginica

inštancia



atribút

Prečo sa vôbec zaoberáme strojovým učením?

6

- Množstvo dát
- Počítače sú lacné, ľudia sú drahí (áno aj bakalári)

Typy úloh SU

Učenie s učiteľom

- Označkové tréningové dáta
- Majorita úloh z praxe
- Cieľom je nájsť dostatočne všeobecnú funkciu $Y=f(X)$

Učenie bez učiteľa

- Žiadne označkové tréningové dáta
- Menšie vstupné náklady (netreba učiteľa)
- Cieľom je nájsť prirodzenú štruktúru v dátach

Učenie s čiastočnou podporou učiteľa

- Malý počet označkových tréningových dát, kombinovaných s väčším počtom neoznačkových dát
- Reálny scenár ak nemáte dost učiteľov

Učenie s učiteľom

- Označované trénovacie dáta
- Majorita úloh z praxe
- Cieľom je nájsť dostatočne všeobecnú funkciu $Y=f(X)$
- Úlohy
 - **klasifikácia** - odhad kategorického Y (spam/pošta)
 - **regresia** - odhad reálneho Y (cena, váha)

Učenie bez učiteľa

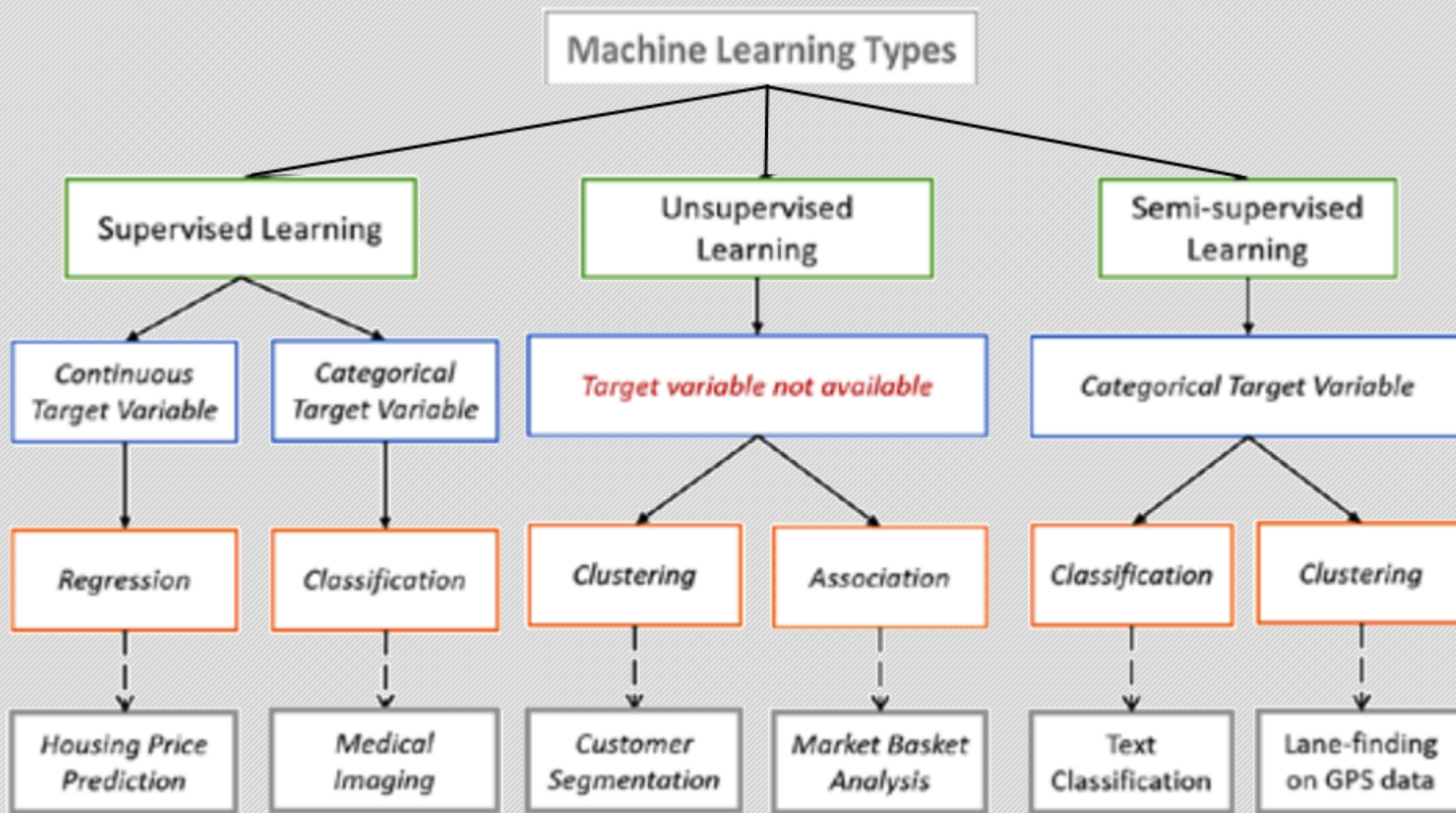
- Žiadne označované trénovacie dáta
- Menšie vstupné náklady (netreba učiteľa)
- Cieľom je nájsť prirodzenú štruktúru v dátach
- Úlohy
 - **zhlukovanie** - segmentácia zákazníkov, častí obrázkov, hľadanie komunit na soc. sieťach
 - **hľadanie asociácií** - nákupný košík, spoločne kupované položky

Učenie s čiastočnou podporou učiteľa

- Malý počet označovaných trénovacích dát, kombinovaných s väčším počtom neoznačovaných dát
- Reálny scenár ak nemáte dost' učiteľov
- Úlohy:
 - **odhad klasifikácie** - časť pozitívnych dát je označených a zvyšku neviete povedať aké sú
 - **odhad zhlukov** - len časť dát je označovaných a vy očakávate zhluky

Rozdelenie úloh strojového učenia III.

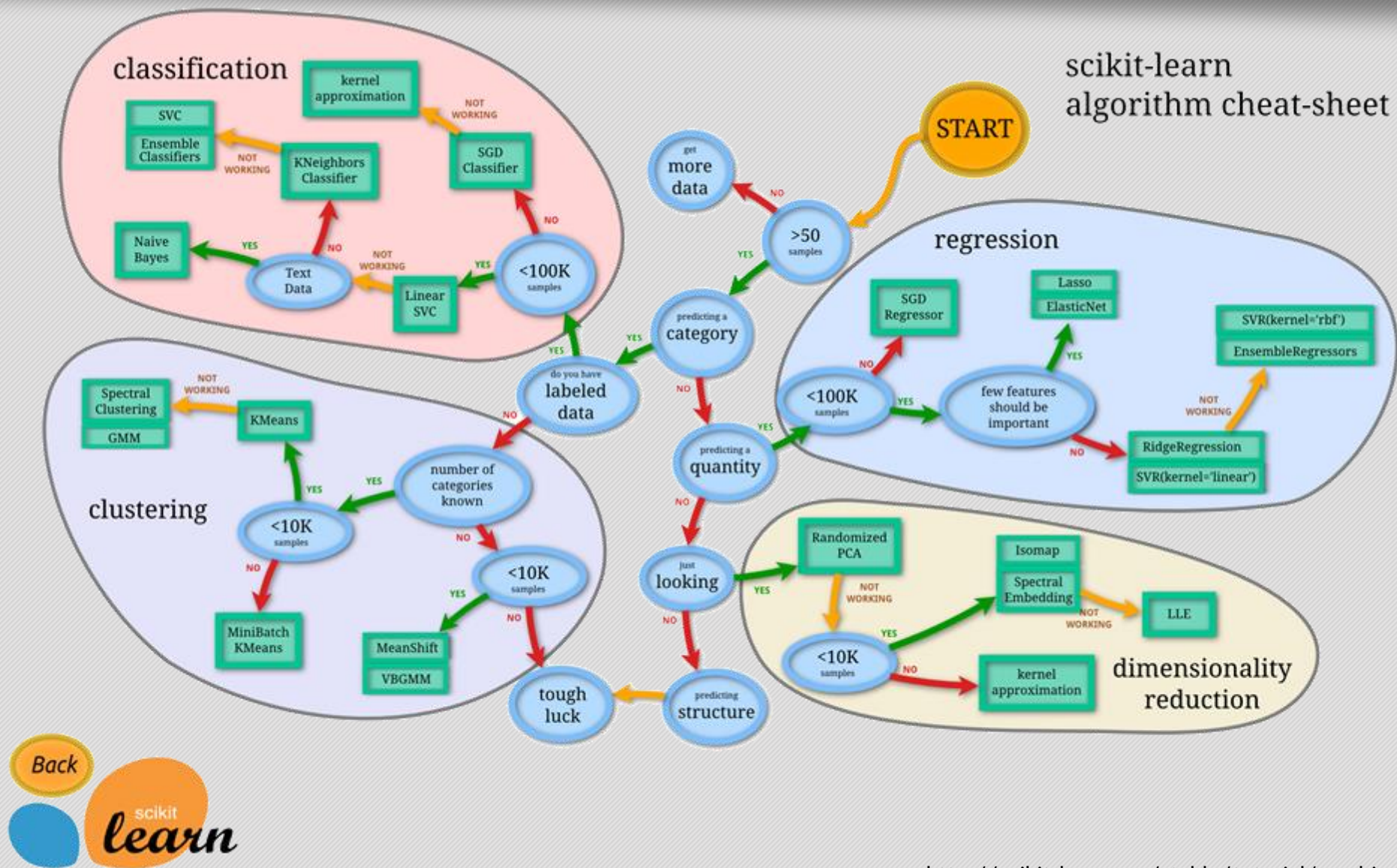
10



- Prístup
- Typ závislej premennej
- Úloha
- Príklad využitia

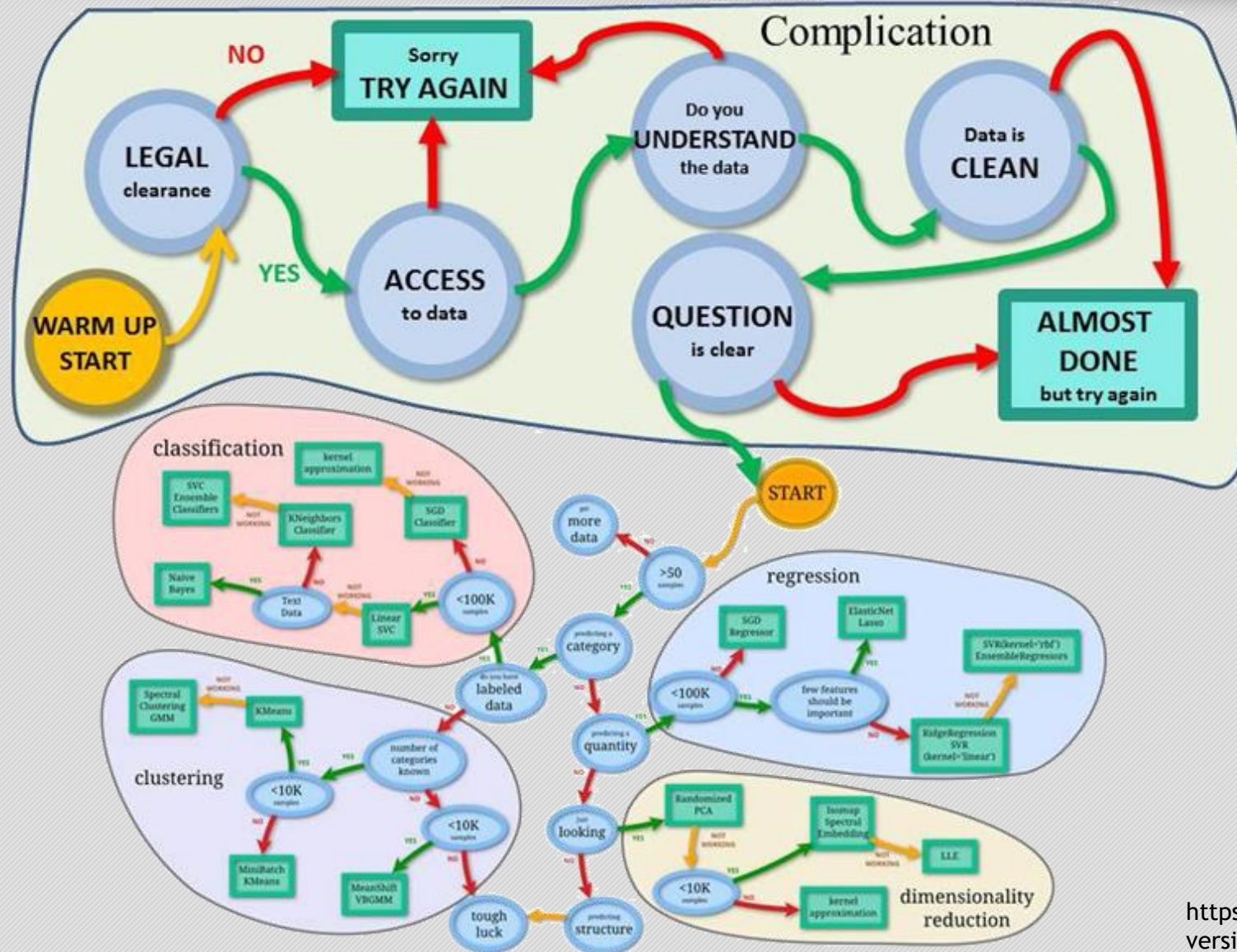
Akou metódou SU mám svoju úlohu riešiť?

11



Skôr ako začnete...

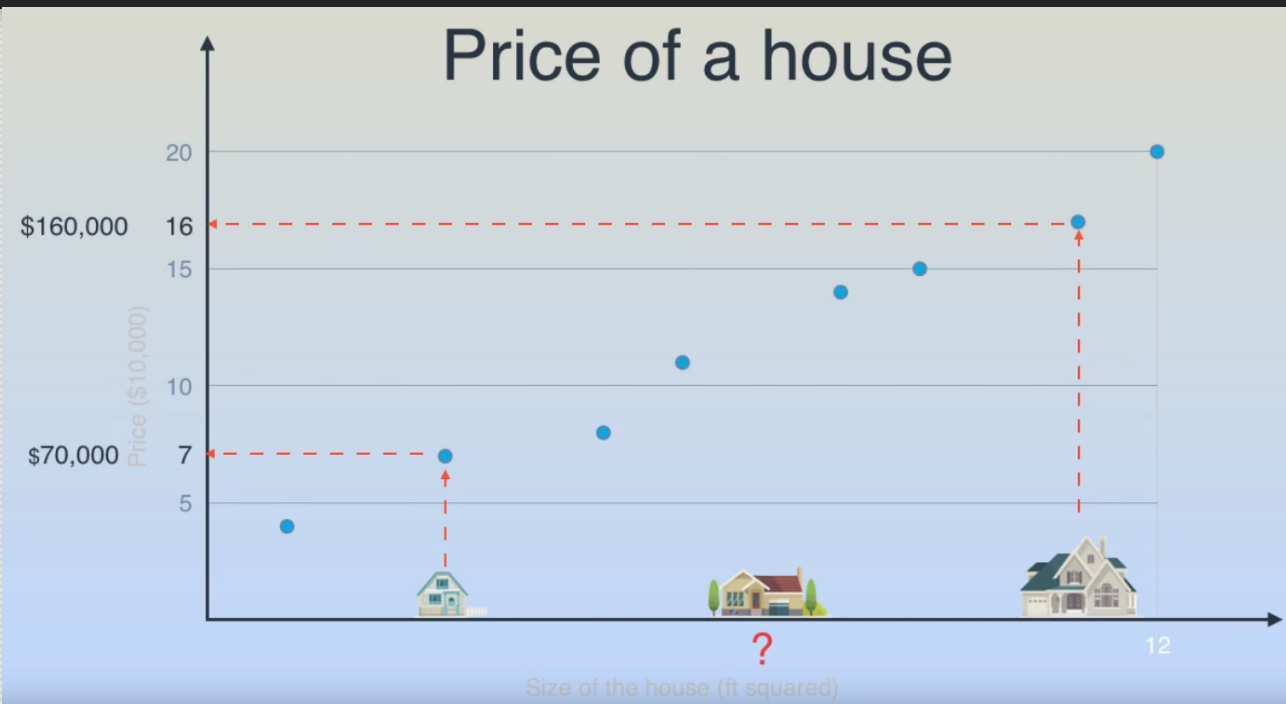
12



Základné metódy riešenia jednotlivých úloh SU

Regresia: Lineárna regresia

14



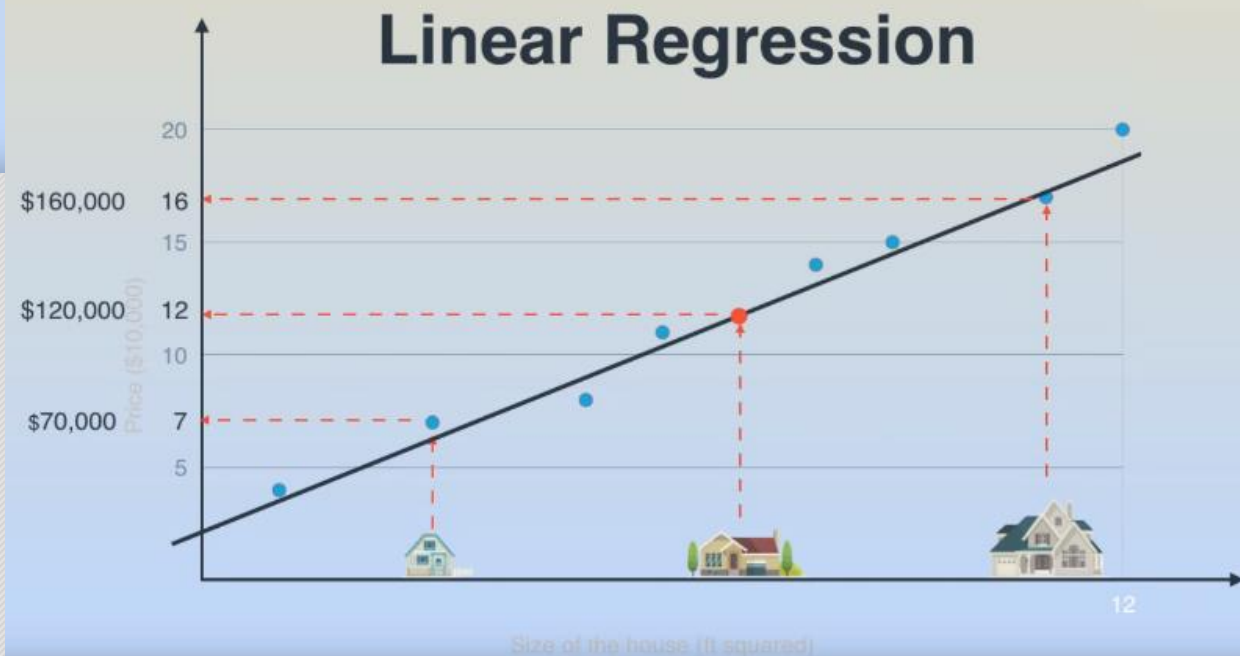
- Ďalšie:

- Polynomial, Ridge, Lasso, Bayesian, Quantile, LAD

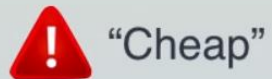
(<http://www.datasciencecentral.com/profiles/blogs/10-types-of-regressions-which-one-to-use>)

A Friendly Introduction to Machine Learning
<https://www.youtube.com/watch?v=lpGxLWOIZy4>

- $Y = f(X)$
- $Y = a + bX$
- $\text{Price} = a + b \cdot \text{Size}$

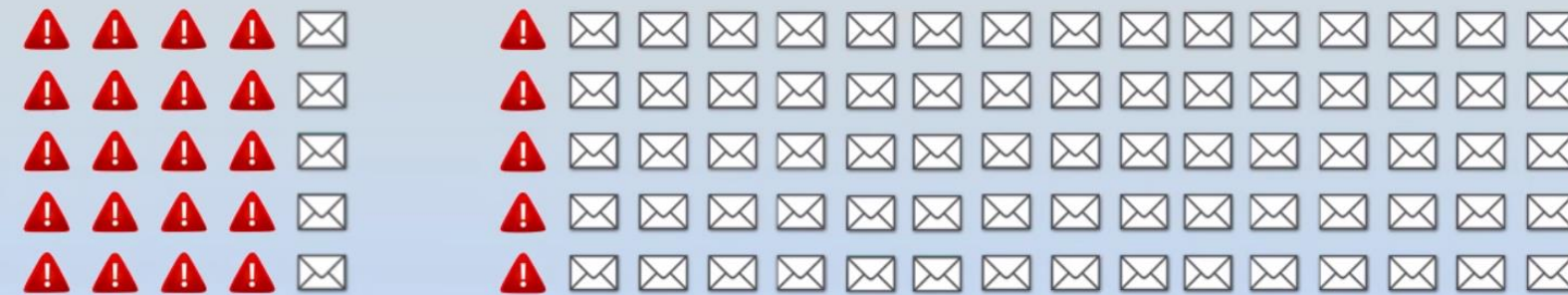


Detecting Spam e-mails



Spam

Non-spam



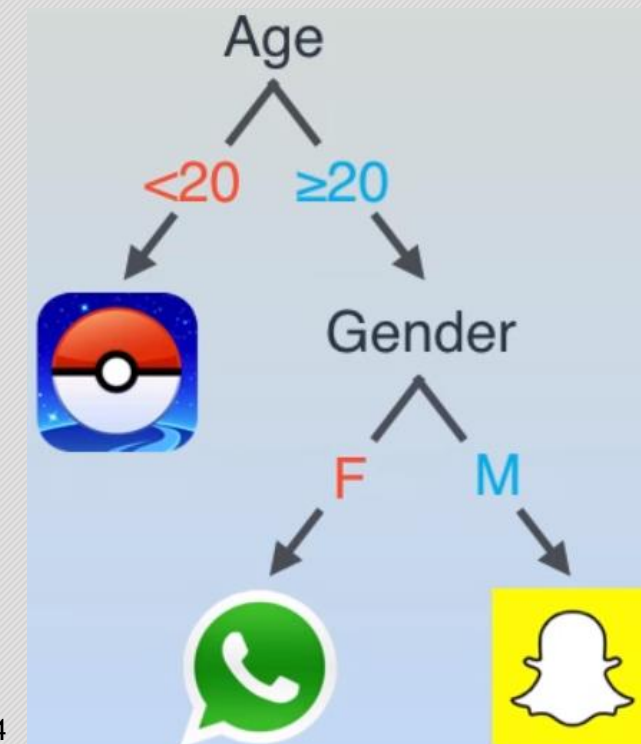
- $Y - \{\text{Spam, Non-spam}\}$
- X
 - obsahuje slovo „Cheap“?
 - má mail predmet?
 - Je písané kapitálkami?
 - ...

Klasifikácia: Rozhodovací strom

16

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	

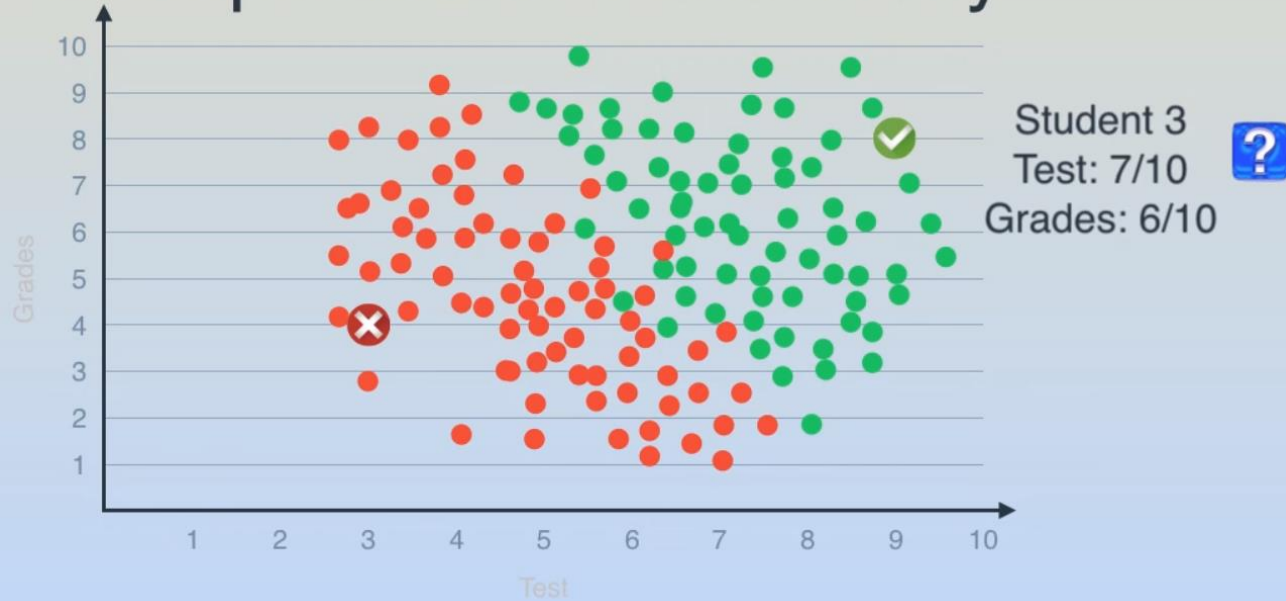
- Postupne vyberáme atribúty s najvyššou rozhodovacou mierou
 - Vek
 - Pohlavie



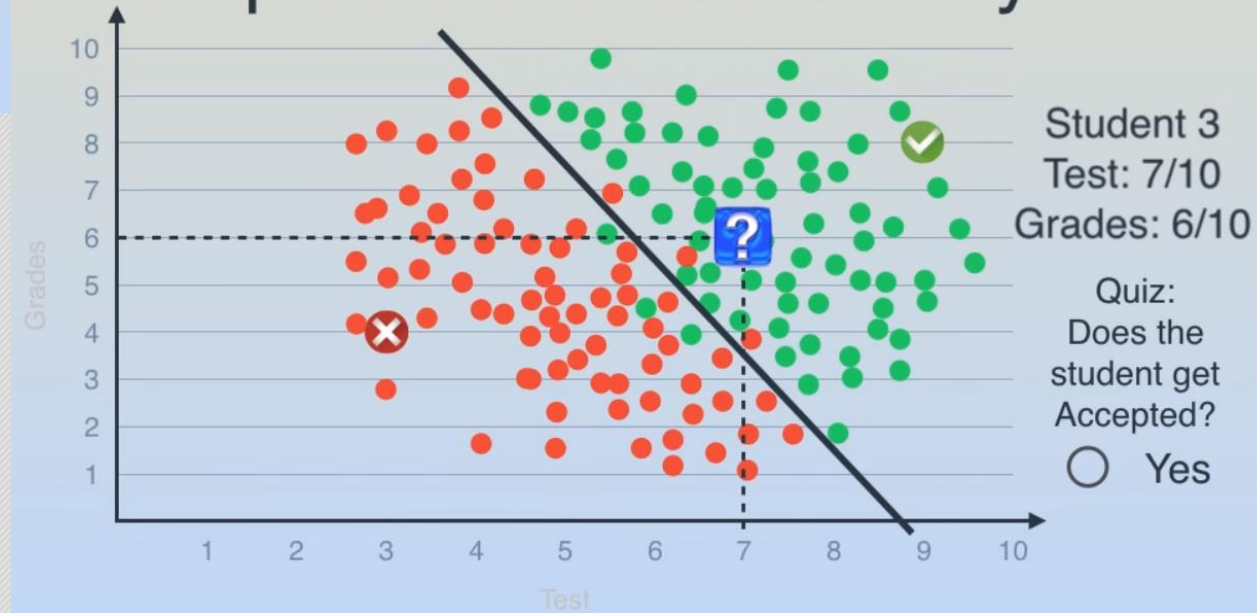
Klasifikácia: SVM I.

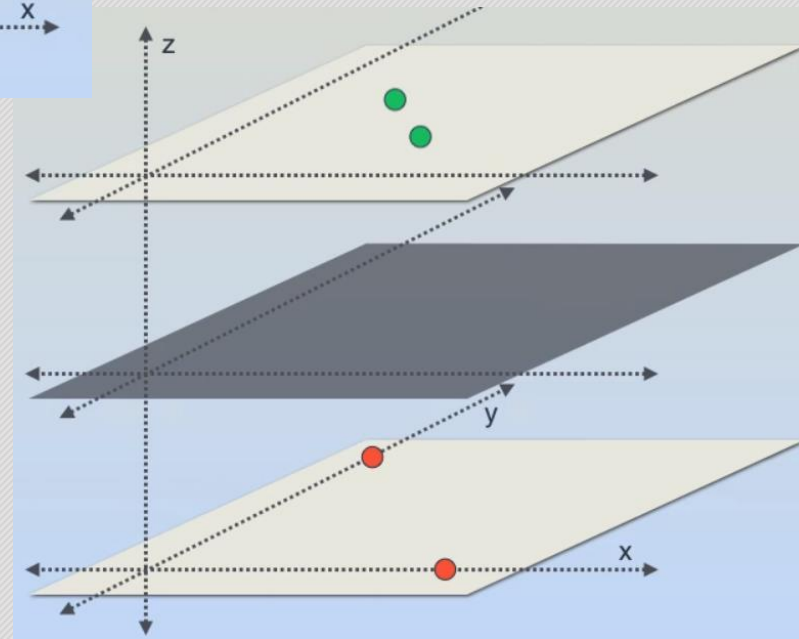
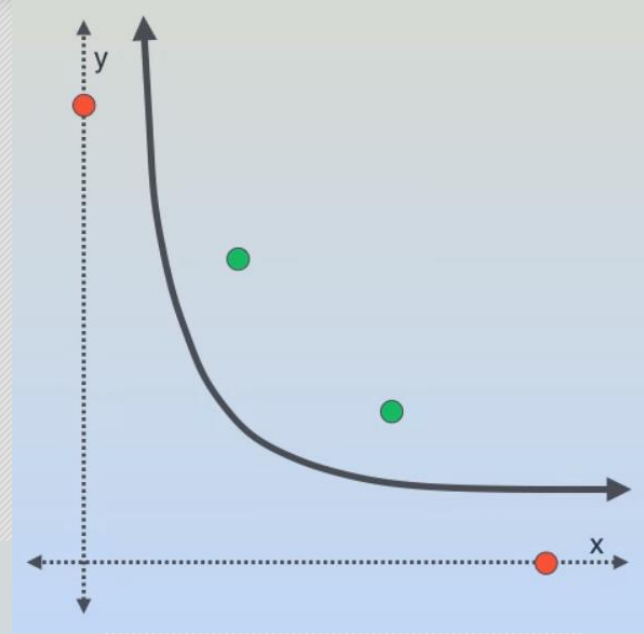
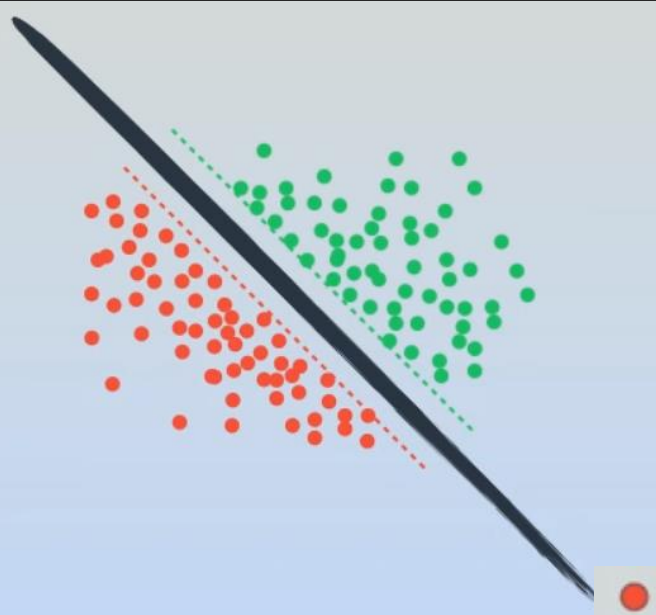
17

Acceptance at a University

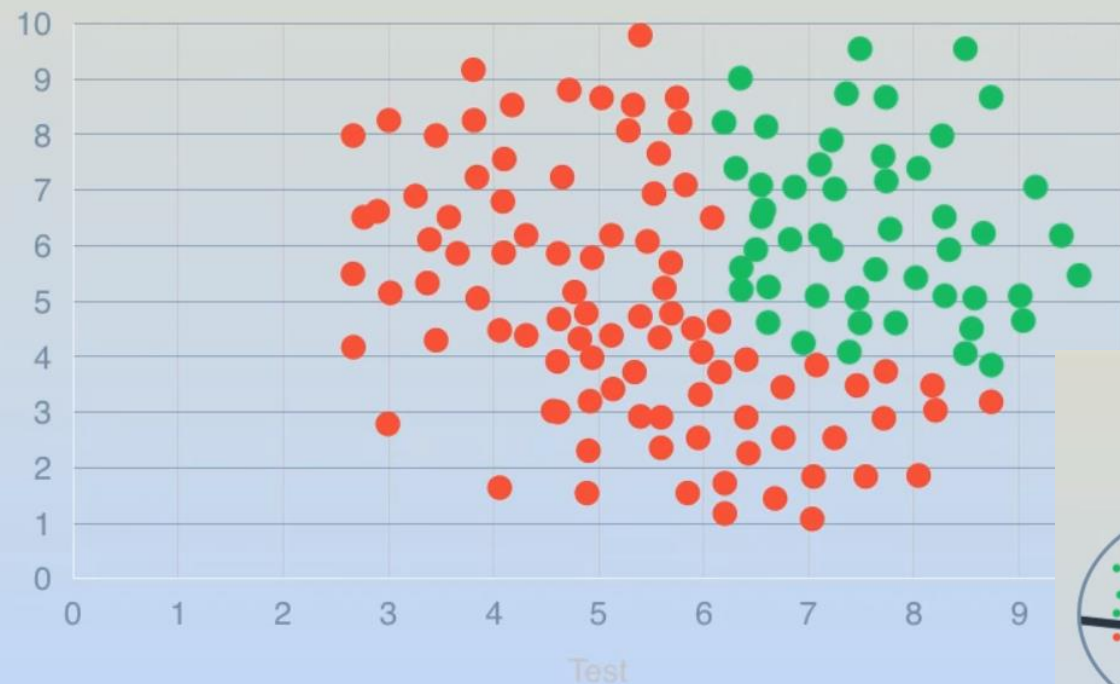


Acceptance at a University

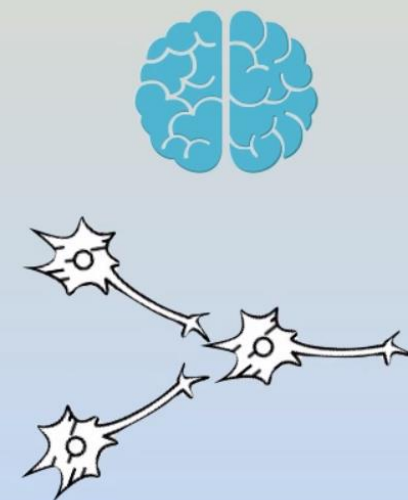
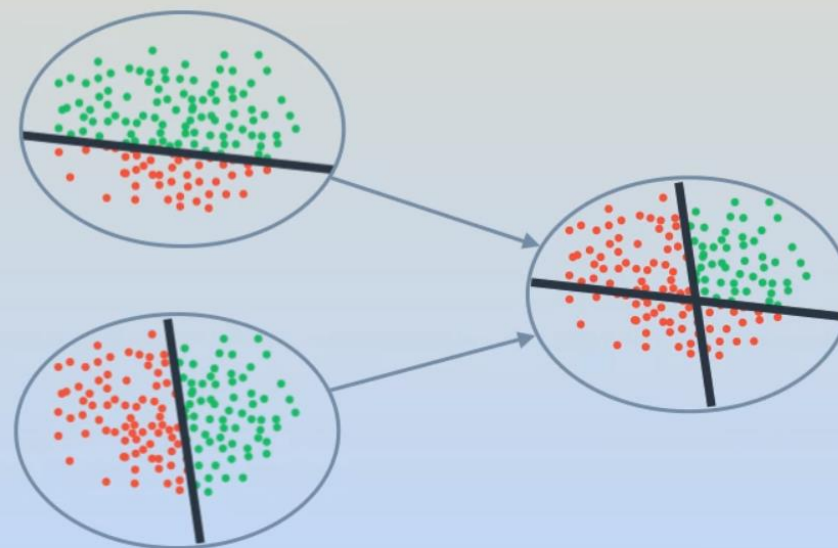




Acceptance at a University

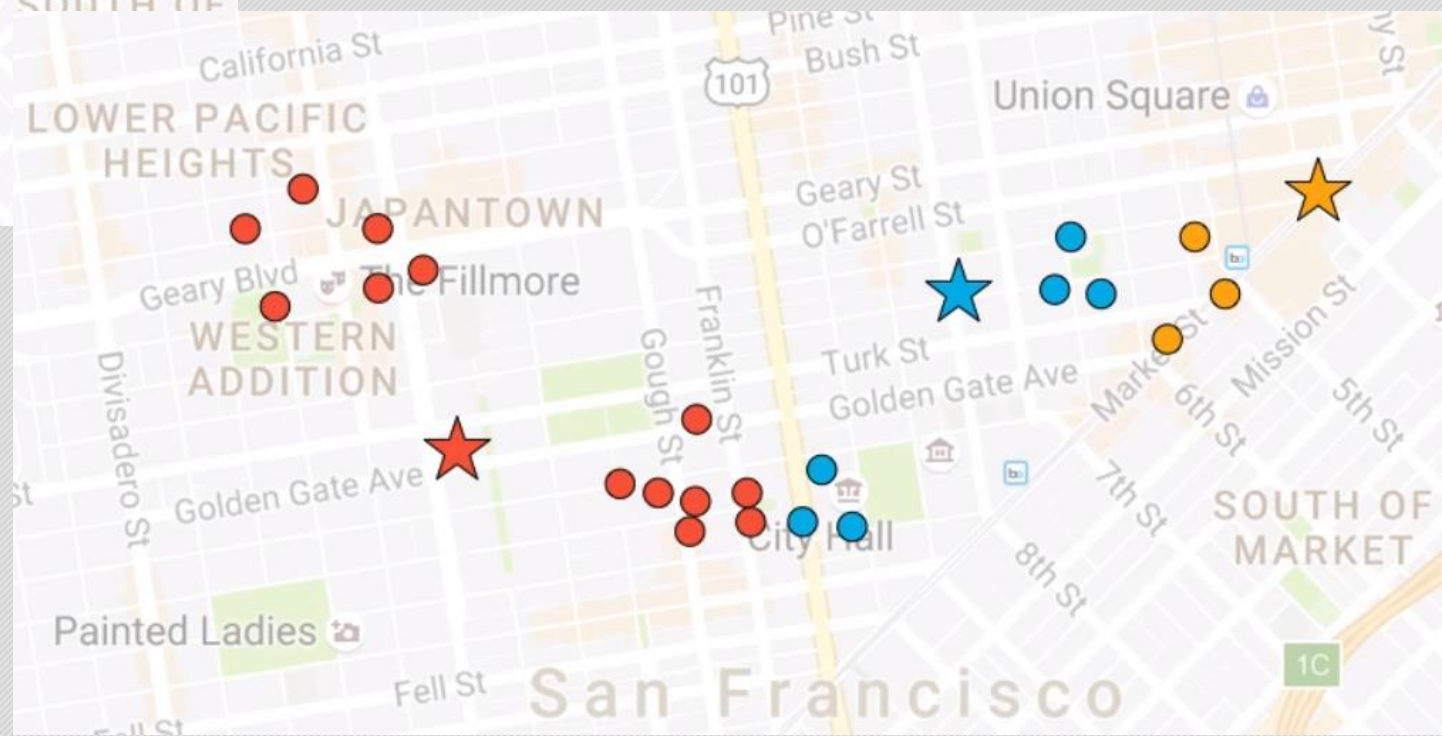
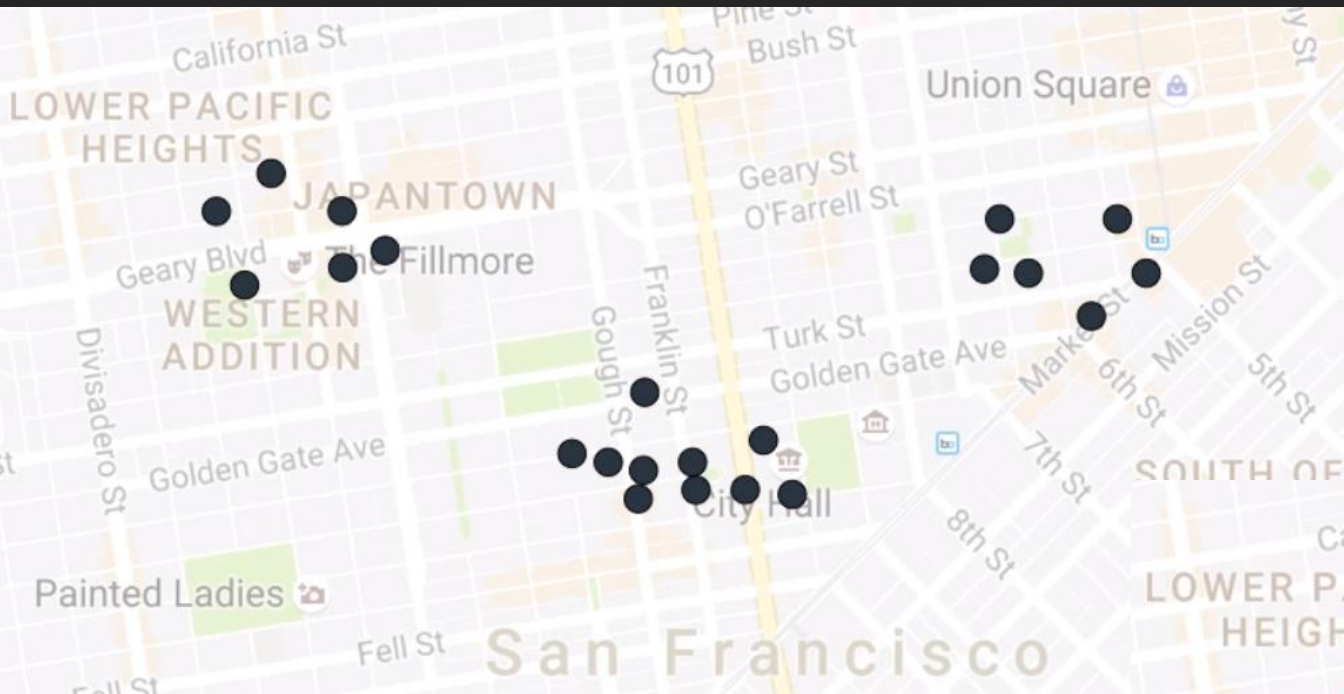


Neural Network



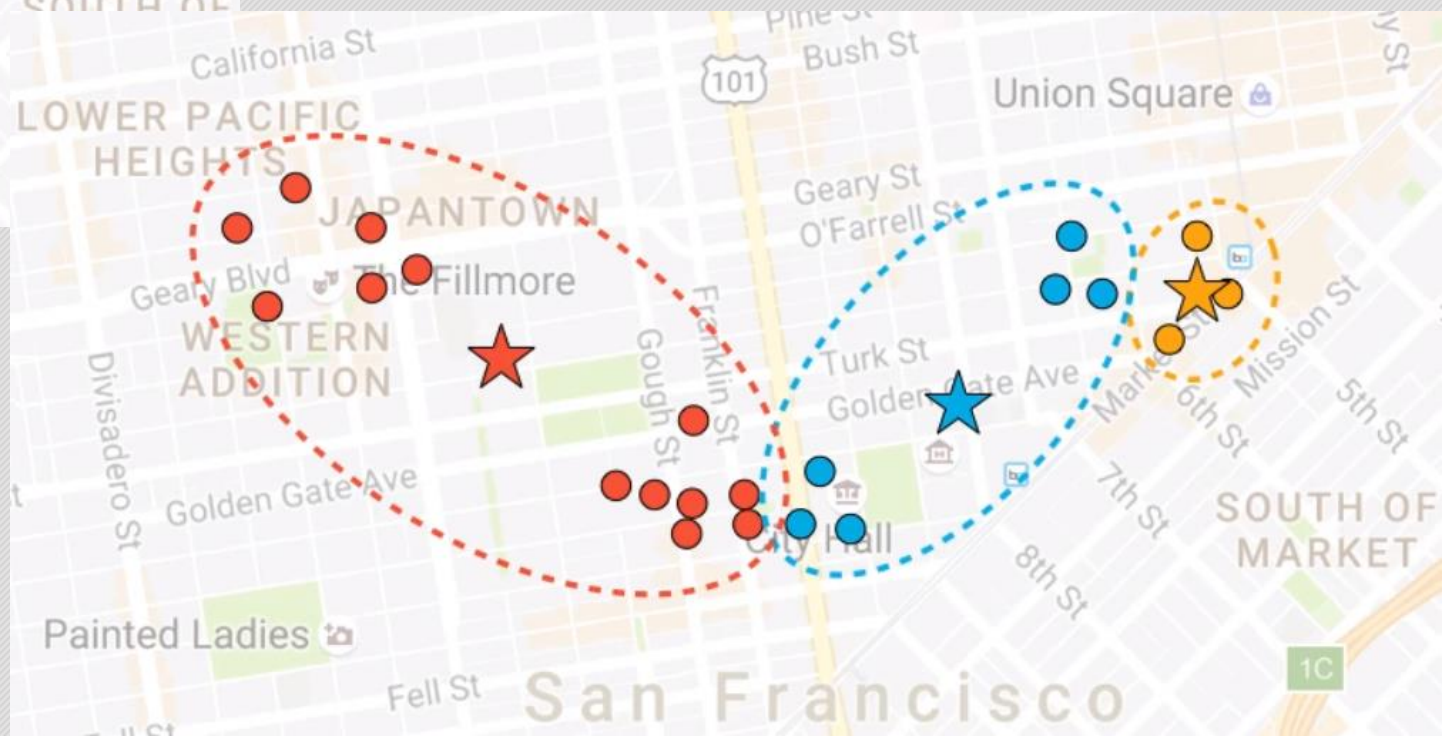
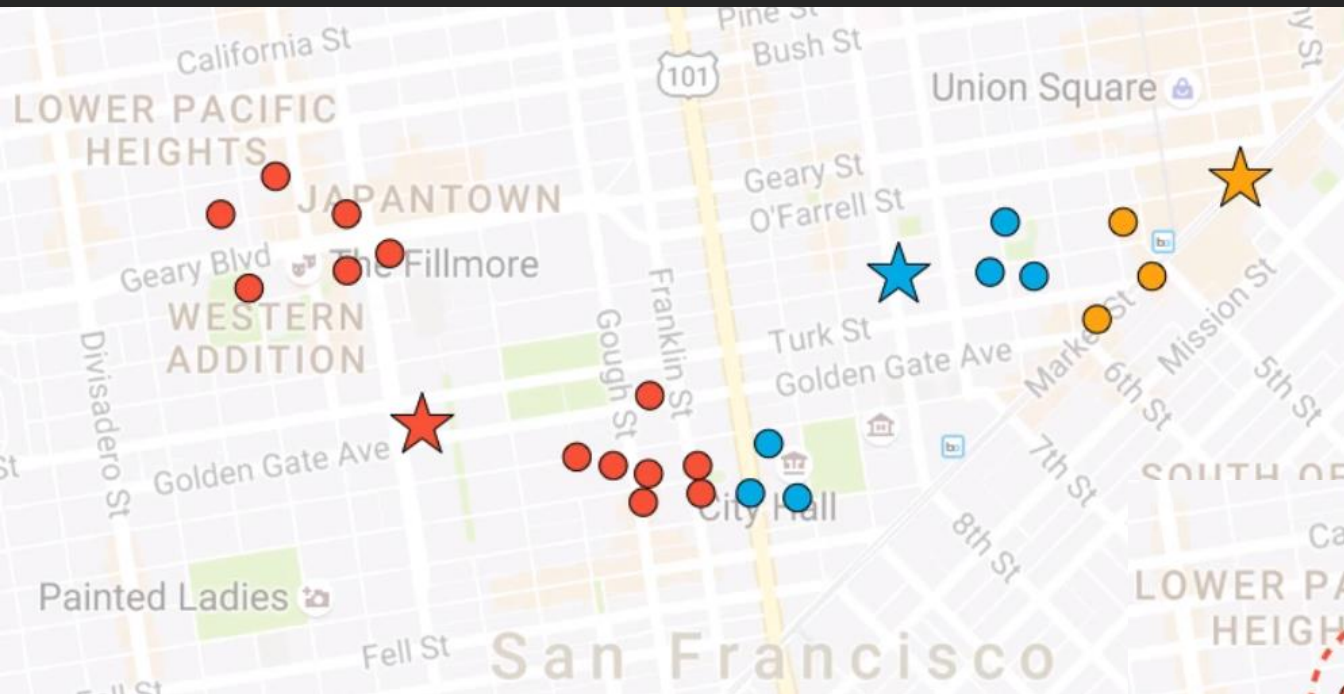
Zhlukovanie: K-Means I.

20



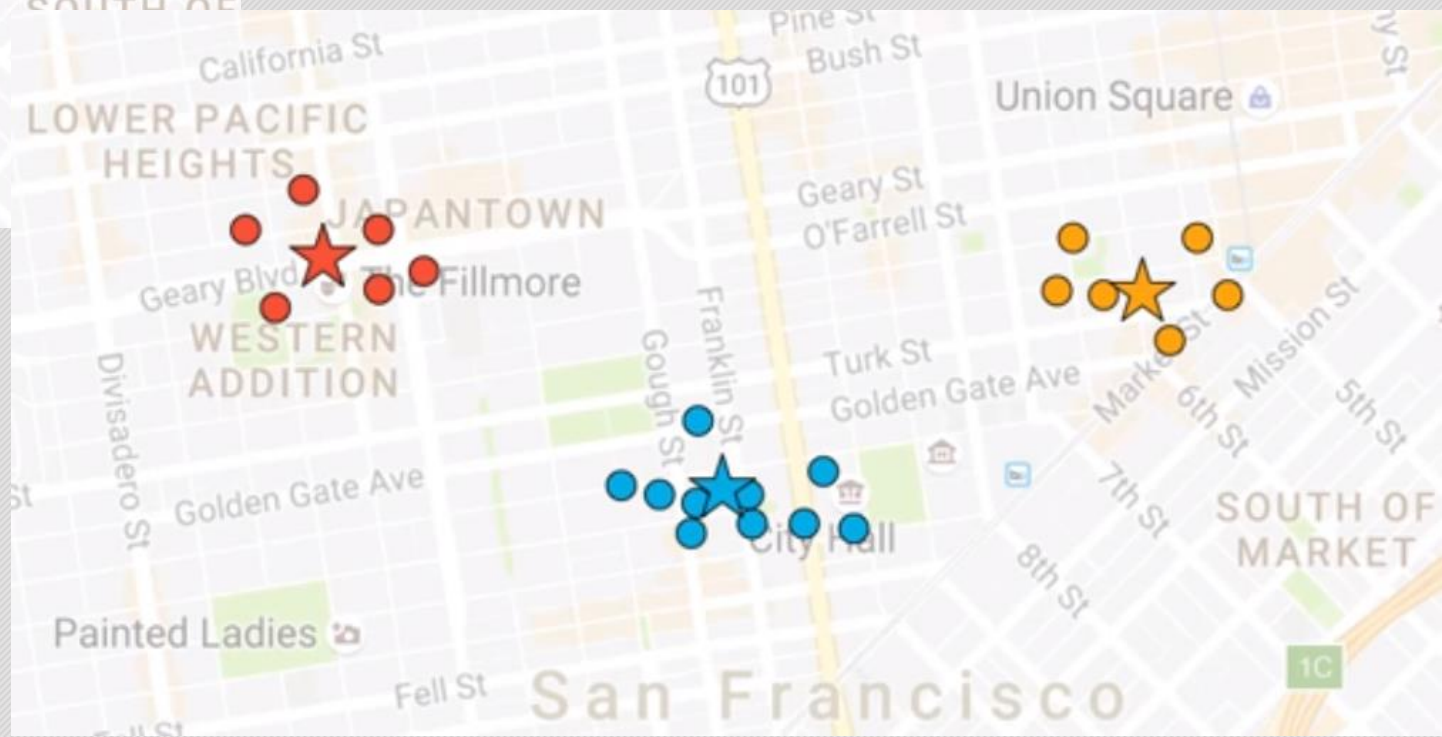
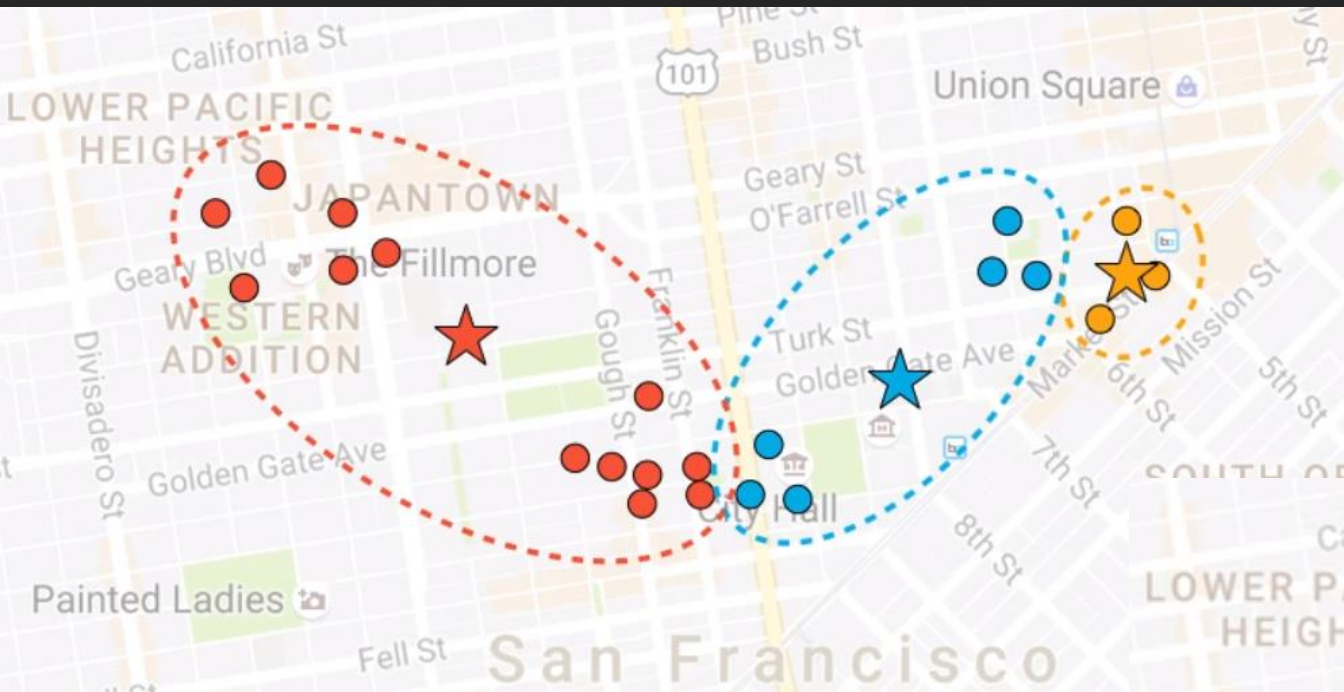
Zhlukovanie: K-Means II.

21



Zhlukovanie: K-Means III.

22



- Urobte si rýchly prehľad oblasti
 - aby ste rozumeli základným pojmom
 - a vedeli čo máte hľadať
- Vybranú oblasť si naštudujte podrobnejšie
 - Sledujte čo používajú iní
 - Čo sa pre daný problém osvedčilo a čo nie
 - Experimentujte
- <https://www.coursera.org/learn/machine-learning>