# Uvod u Mašinsko učenje

Milan M.Milosavljević

# Neke izjava o Mašinskom učenju

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, CEO, Yahoo)

# Šta je mašinsko učenje

- Automatizacija programiranja
- Omogućava samoprogramiranje
- Pisanje softvera je usko grlo računarstva
- Neka sami podaci „rade posao“ umesto klasičnog programiranja

## Tradicionalno Programiranje


## Mašinsko učenje

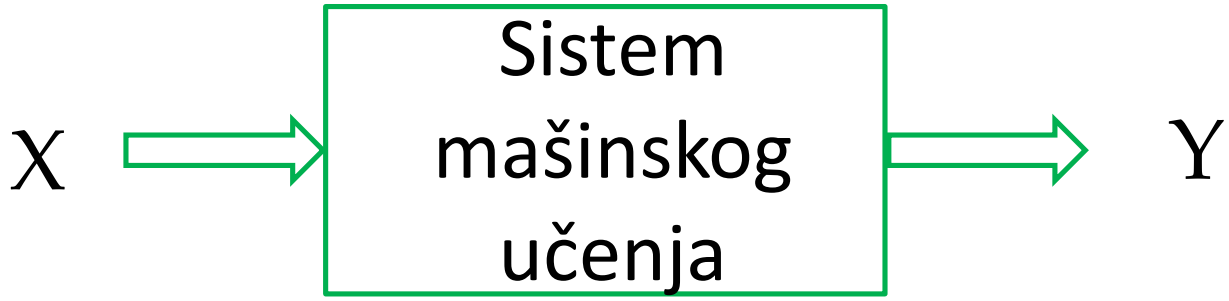
# Analogija sa baštovanstvom

- Seme = Algoritmi
- Zemljište = Podaci
- Baštovan = Vi
- Rastinje = Programi

# Vrste mašinskog učenja

- Induktivno učenje (Supervised learning)
  - Podaci za obučavanje sadrža željeni izlaz
- Samoobučavanje (Unsupervised learning)
  - Podaci za obučavanje ne sadrže željeni izlaz
- Semi-induktivno učenje
  - Pored označenih primera, dominiraju neoznačeni primeri
- Obučavanje sa pojačavanjem (Reinforcement learning)
  - Nagradjivanje sekvence akcija

# Klasifikacija vs regresija




$X$ – ulaz: vektor beležja

$Y$ – izlaz: nominalan (klasifikacija)
kontinualan (regresija)

# ZAŠTO MAŠINSKO UČENJE ?

- Mašinsko učenje je u užem smislu programiranje računarskih mašina u cilju optimizacije pogodnog kriterijuma optimalnosti, na osnovu raspoloživih podataka ili prošlih iskustava.
- Nema potrebe da učimo sračunavanje korena kvadratne jednačine
- Učenje je neophodno kada:
  - Ne postoje eksperti za datu oblast (navigacija na Marsu, dijagnoza na osnovu genskih eksprsija),
  - Ljudi nisu u stanju da objasne svoju ekspertizu (prepoznavanje govora, prepoznavanje znakova)
  - Rešenje se menja u vremenu (rutiranje u računarskim mrežama)
  - Rešenje je neophodno prilagoditi konkretnim situacijama i slučajevima ( biometrija)

# O čemu se radi kada govorimo o mašinskom učenju

- Učenje modela na osnovu podataka u vezi sa pojedinačnim primerima
- Danas su podaci jeftini, dostupni i raspoloživi u velikim količinama (data warehouses, data marts); znanje je skupo i retko.
- Primer maloprodaje: Iz kupovnih transakcija se uči ponašanje kupaca:

  *Ko kupuje pelene, često kupuje i pivo*
- Učenje modela koji je dobra i korisna aproksimacija podataka.

# Data Mining

- Maloprodaja: Market basket analiza, Customer relationship management (CRM)
- Finansije: kreditni skorovi, detekcija prevara
- Proizvodnja: upravljanje, robotika, pronalaženje kvarova
- Medicina: medicinska dijagnostika
- Telekomunikacije: spam filteri, detekcija upada u računarske mreže
- Bioinformatika: motifs, alignment
- Web mining: Search engines
- ...

# Šta je mašinsko učenje?

- Optimizacija kriterijuma performanse na osnovu primera ili prošlih iskustava
- Uloga statistike: zaključivanje na osnovu uzoraka
- Uloga računarskih nauka: efikasni algoritmi za
  - rešavanje optimizacionih problema
  - reprezentacija i evaluacija modela u cilju zaključivanja

# Primene

- Asocijacije
- Obučavanje sa učiteljem (Supervised Learning)
  - Klasifikacija
  - Regresija
- Samoobučavanje (Unsupervised Learning)
- Učenje sa pojačavanjem (Reinforcement Learning)

# Učenje asocijacija

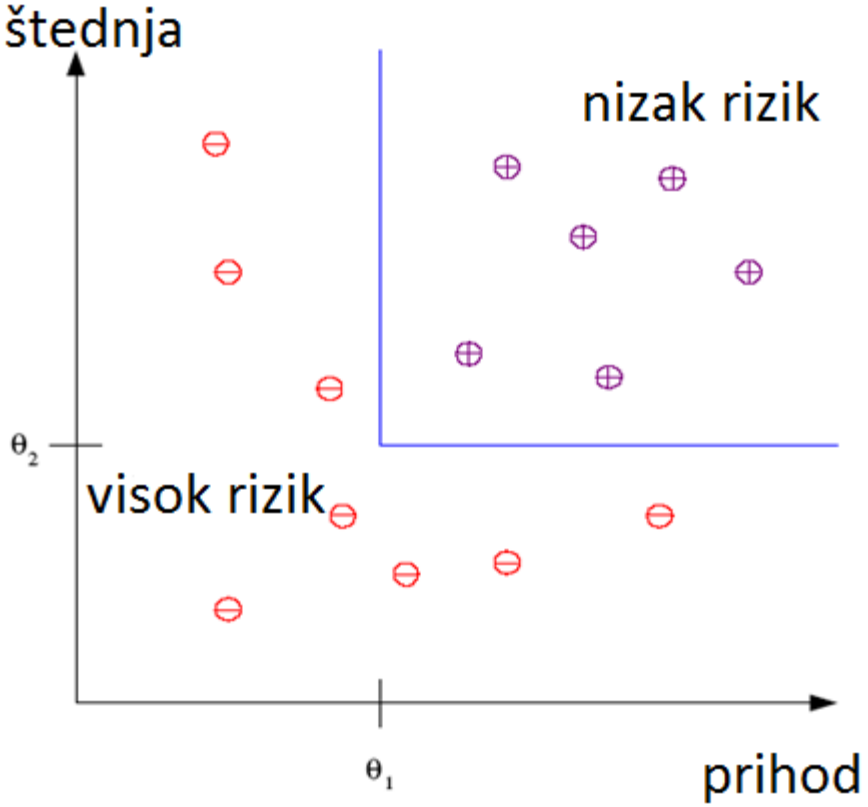
- Analiza sadržaja korpe pri kupovini:

  $P(Y \mid X)$ verovatnoća da neko ko je kupio $X$ takodje kupi i $Y$, gde su $X$ i $Y$ proizvodi ili servisi.

  Primer: $P(\text{čips} \mid \text{pivo}) = 0.7$

# Klasifikacija

- Primer: Kreditni skorovi
- Razlikovanje niskorizičnih i visokorizičnih klijenata na osnovu njihovog prihoda i visine štednih uloga




Disckriminacija: IF *prihod* > $\theta_1$ AND *štednja* > $\theta_2$

THEN nizak rizik ELSE visok rizik

# Klasifikacija: Primene

- Prepoznavanje oblika (Pattern recognition)
- Prepoznavanje lica: poza, osvetljenje, okluzija (naočare, brada), šminka, frizura
- Prepoznavanje znakova: štampani, rukom pisani
- Prepoznavanje govora: vremenske medjuzavisnosti parametara govora.
- Medicinska dijagnostika: od simptoma ka bolestima
- Biometrija: identifikacija/autentifikacija pomoću fizičkih karakteristika ili ponašanja: lice, iris, potpis, otisak prstiju, način hoda,...
- ...

# Prepoznavanje lica

Obučavajući skup primera za jednu osobu

Test slike lica

ORL dataset,
AT&T Laboratories,
Cambridge UK

# Regresija

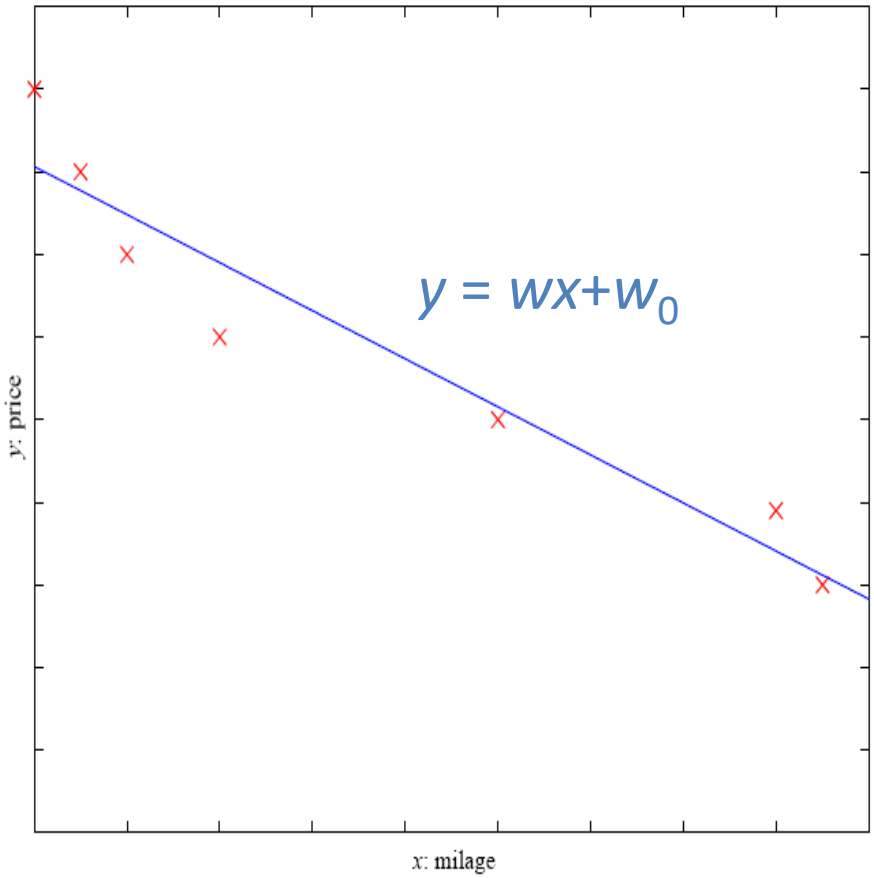
- Primer: cena polovnih automobila
- $x$ : svojstva automobila

  $y$ : cena

  $y = g(x \mid \theta)$

  $g(\ )$ model,

  $\theta$ parameteri



# Obučavanje sa učiteljem: upotreba

- Predikcija budućih slučajeva: korišćenje naučenih pravila u cilju predikcije budućih izlaza na osnovu budućih ulaza
- Ekstrakcija znanja: pravila se lako razumeju
- Kompresija: pravila su jednostavnija od podataka koje objašnjavaju
- Detekcija Outlier-a: izuzeci koji nisu pokriveni naučenim pravilima ( detekcija prevara)

# Samoobučavanje

- Učenje koncepta “Šta se normalno dešava”
- U toku učenja nije prisutan izlaz
- Klasterovanje: grupisanje sličnih primera
- Primeri primene
  - segmentacija korisnika u CRM
  - kompresija slike: kvantizacija boja
  - bioinformatika: učenje motifs-a

# Učenje sa pojačavanjem

- Učenje strategije:  sekvence izlaza
- Nije na raspolaganju informacija o izlazu, već zakašnjena nagrada
- Credit assignment problem
- Igre
- Robot u lavirintu
- Višestruki agenti u parcijalno observabilnom okruženju ...

# Resursi na internetu: podaci

- UCI Repository: http://www.ics.uci.edu/~mlearn/MLRepository.html
- UCI KDD Archive: http://kdd.ics.uci.edu/summary.data.application.html
- Statlib: http://lib.stat.cmu.edu/
- Delve: http://www.cs.utoronto.ca/~delve/