

Mašinsko učenje

Uvod u mašinsko učenje

Nemanja Maček

- Veštačka inteligencija i mašinsko učenje
- Induktivno empirijsko učenje funkcionalnih preslikavanja
- Generalizacija, podela na tri skupa i problem prenaučivosti
- Klasifikacija

Šta je veštačka inteligencija?

- VI trenutno obuhvata mnoštvo podoblasti, od vrlo opštih, kao što su učenje i percepcija, do uskih zadataka kao što je:
 - igranje šaha,
 - dokazivanje matematičkih teorema,
 - medicinska dijagnostika,
 - automatsko prevođenje,
 - prepoznavanje govora,
 - robotika, itd.

Šta je veštačka inteligencija?

- I pored poluvekovne istorije VI je i dalje oblast koju je teško precizno definisati.
- Primeri nekih definicija dobijenih dihotomijom u odnosu na dimenzije rezonovanja.
 - Sistemi koji razmišljaju kao ljudi (ljudsko rezonovanje).
 - “[Automatizacija] aktivnosti koje asociramo sa ljudskim razmišljanjem, aktivnosti kao što su donošenje odluka, rešavanje problema, učenje...”
 - Sistemi koji razmišljaju racionalno (racionalno rezonovanje).
 - “Izučavanje računanja koje omogućava opažanje, rasuđivanje i delovanje.”
 - Sistemi koji se ponašaju kao ljudi (ljudsko ponašanje).
 - “Izučavanje kako naterati računare da rade stvari u kojima su, trenutno, ljudi bolji.”
 - Sistemi koji se ponašaju racionalno (racionalno ponašanje).
 - “Računarska inteligencija je nauka o dizajniranju inteligentnih agenata.”
 - “VI ... se bavi inteligentnim ponašanjem veštackih naprava.”

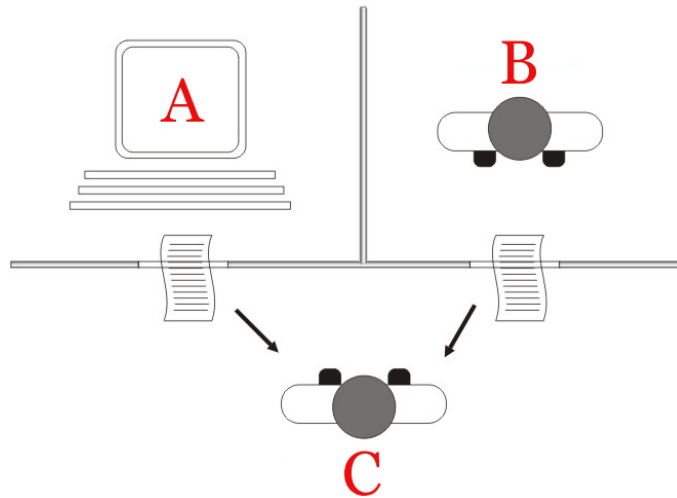
Šta je veštačka inteligencija?

- Ljudski-centrirani pristupi dominantno moraju biti empirijska nauka.
 - Uključuje hipotezu i eksperimentalnu potvrdu o ljudskom ponašanju.
- Racionalistički pristup koristi kombinaciju matematike i inženjerstva.

Veštačka inteligencija i mašinsko učenje

Pristup zasnovan na Tjuringovom testu.

- Alan Tjuring (1950): “*Computing Machinery and Intelligence*”.
 - Umesto pitanja da li mašine mogu da misle, postavlja se pitanje da li mašina može da prodje test inteligentnog ponašanja.
 - Test je nakon toga nazvan Tjuringov test.



Pristup zasnovan na Tjuringovom testu.

- Mašina prolazi Tjuringov test ako osoba ispitivač, posle postavljanja nekoliko pitanja u pisanoj formi, ne može da odredi da li je pisani odgovor dao čovek ili mašina.
- Mašina koja bi eventualno prošla Tjuringov test, morala bi da ima sledeće sposobnosti:
 - obrada prirodnih jezika (da bi mogla uspešno da komunicira u prirodnom jeziku),
 - reprezentacija znanja (da bi memorisala ono što zna i prima na osnovu senzora),
 - automatsko rezonovanje (da bi koristila memorisane informacije za odgovaranje na pitanja i za donošenje novih zaključaka) i
 - mašinsko učenje (da bi se adaptirala novim okolnostima).



Pristup zasnovan na Tjuringovom testu.

- Tjuringov test je namerno izbegavao direktnu fizičku interakciju između ispitivača i računara.
- Razlog: fizička simulacija čoveka nepotrebna za inteligenciju.
- Takozvani potpuni Tjuringov test uključuje video signal.
 - Ispitivač može da testira percepcijske i motorne sposobnosti ispitanika.
- Da bi prošao potpuni Turingov test racunar mora da ovlada:
 - računarskom vizijom (u cilju vizuelene percepcije okoline i objekata u njoj)
 - robotikom (radi kretanja u prostoru i manipulacijom objekata u njemu).
- Ovih 6 disciplina cine dominantan deo VI!



10 hottest AI technologies based on Forrester's analysis.

- Natural Language Generation: Producing text from computer data.
 - Currently used in customer service, report generation, and summarizing business intelligence insights.
 - Sample vendors: Attivio, Automated Insights, Cambridge Semantics, Digital Reasoning, Lucidworks, Narrative Science, SAS, Yseop.
- Speech Recognition: Transcribe and transform human speech into format useful for computer applications.
 - Currently used in interactive voice response systems and mobile applications.
 - Sample vendors: NICE, Nuance Communications, OpenText, Verint Systems.

10 hottest AI technologies based on Forrester's analysis.

- Virtual Agents: From simple chatbots to advanced systems that can network with humans.
 - Currently used in customer service and as a smart home manager.
 - Sample vendors: Amazon, Apple, Artificial Solutions, Assist AI, Creative Virtual, Google, IBM, IPsoft, Microsoft, Satisfi.
- Machine Learning Platforms: Providing algorithms, APIs, development and training toolkits, data, as well as computing power to design, train, and deploy models into applications, processes, and other machines.
 - Currently used in a wide range of enterprise applications, mostly involving prediction or classification.
 - Sample vendors: Amazon, Fractal Analytics, Google, H2O.ai, Microsoft, SAS, Skytree.

10 hottest AI technologies based on Forrester's analysis.

- AI-optimized Hardware: Graphics processing units (GPU) and appliances specifically designed and architected to efficiently run AI-oriented computational jobs.
 - Currently primarily making a difference in deep learning applications.
 - Sample vendors: Alluviate, Cray, Google, IBM, Intel, Nvidia.
- Decision Management: Engines that insert rules and logic into AI systems and used for initial setup/training and ongoing maintenance and tuning.
 - A mature technology, it is used in a wide variety of enterprise applications, assisting in or performing automated decision-making.
 - Sample vendors: Advanced Systems Concepts, Informatica, Maana, Pegasystems, UiPath.

10 hottest AI technologies based on Forrester's analysis.

- Deep Learning Platforms: A special type of machine learning consisting of artificial neural networks with multiple abstraction layers.
 - Currently primarily used in pattern recognition and classification applications supported by very large data sets.
 - Sample vendors: Deep Instinct, Ersatz Labs, Fluid AI, MathWorks, Peltarion, Saffron Technology, Sentient Technologies.
- Biometrics: Enable more natural interactions between humans and machines, including but not limited to image and touch recognition, speech, and body language.
 - Currently used primarily in market research.
 - Sample vendors: 3VR, Affectiva, Agnitio, FaceFirst, Sensory, ...

10 hottest AI technologies based on Forrester's analysis.

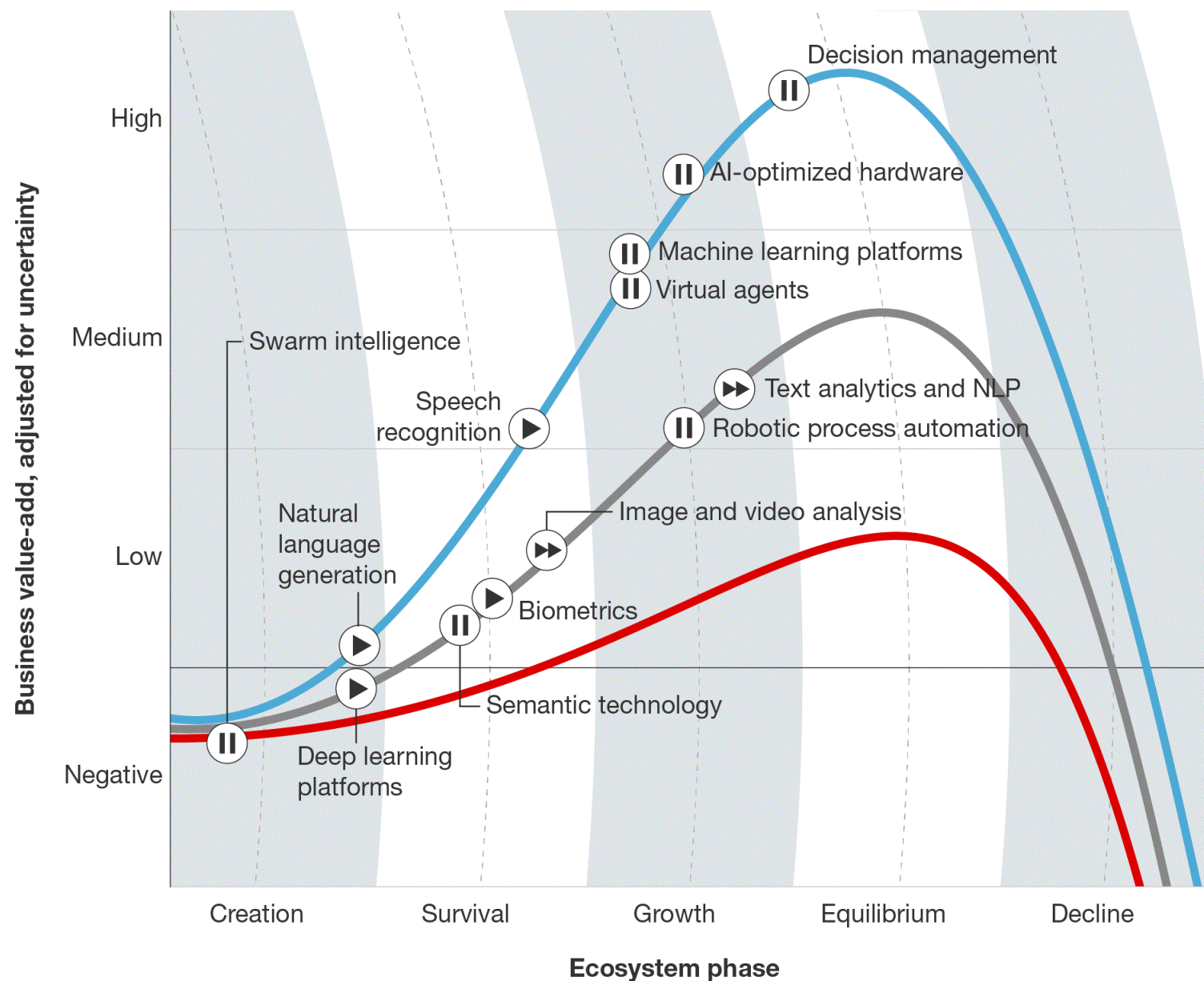
- Robotic Process Automation: Using scripts and other methods to automate human action to support efficient business processes.
 - Currently used where it's too expensive or inefficient for humans to execute a task or a process.
 - Sample vendors: Advanced Systems Concepts, Blue Prism, UiPath, ...
- Natural language processing (NLP): Uses and supports text analytics by facilitating the understanding of sentence structure and meaning, sentiment, and intent through statistical and machine learning methods.
 - Currently used in fraud detection and security, a wide range of automated assistants, and applications for mining unstructured data.
 - Sample vendors: Basis Technology, Coveo, Expert System, Indico, ...

Trajectory:

- Significant success
- Moderate success
- Minimal success

Time to reach next phase:

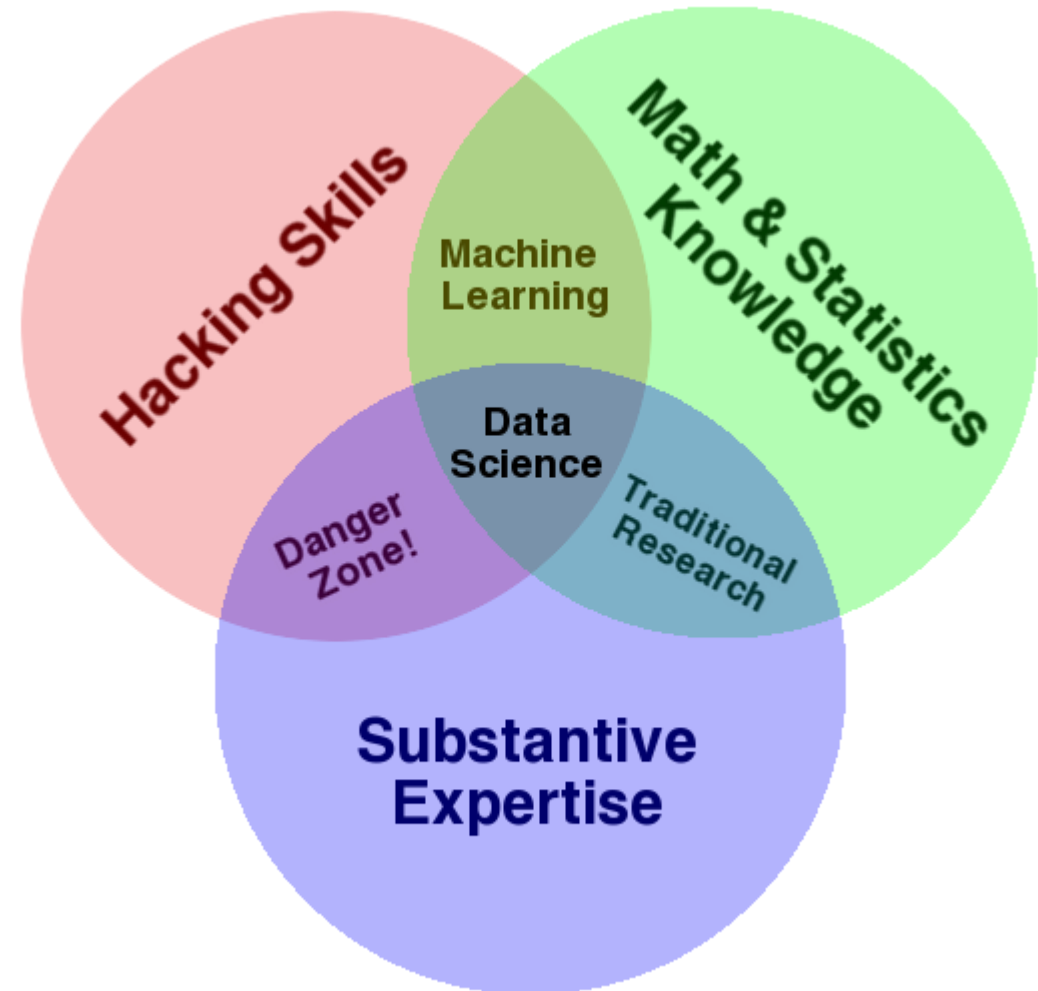
- <1 year
- 1 to 3 years
- 3 to 5 years
- 5 to 10 years
- >10 years



Veštačka inteligencija i mašinsko učenje

Evolucija ka nauci o podacima.

- 1950. Kibernetika.
- 1960. Veštačka inteligencija.
- 1970. Prepoznavanje oblika.
- 1980. Mašinsko učenje.
- 1990. Statistička teorija učenja.
- 2010. Duboko-obučavajući sistemi.
- 2015. Nauka o podacima.



Fenomen podataka.

- “U Boga verujemo, svi ostali neka donesu podatke.” – William E. Deming
 - Na Vebu se ova izjava pripisuje podjednako i Vilijemu Demingu i Robertu Hajdenu.
 - Profesor Hajden tvrdi da ovo nije njegova izjava.
 - Ironija je u tome da se ne mogu naći “podaci” koji bi potvrdili da je Deming zaista njen autor.

W. Edwards Deming



Born	October 14, 1900 Sioux City, Iowa
Died	December 20, 1993 (aged 93) Washington, D.C.
Fields	Statistician
Alma mater	University of Wyoming BS University of Colorado MS Yale University PhD
Influences	Walter A. Shewhart

Šta je učenje?

- Učenje je proces koji se kod ljudi odvija gotovo neprekidno.
- Postoje različite vrste učenja:
 - čisto memorisanje podataka,
 - učenje motornih veština
 - sticanja sposobnosti analitičkog i kreativnog mišljenja, itd.
- Sposobnost učenja se smatra osnovnom odlikom inteligentnih bića!
 - Nije iznenađujuće što je mašinsko učenje jedna od centralnih oblasti istraživanja u VI.
- Mašinsko učenje je oblast koja proučava procese na kojima se zasniva učenje kod ljudi i kod veštačkih sistema.
 - Veštački sistemi koji su sposobni da uče evidentno poboljšavaju svoje performanse.
 - Biološki sistemi povećavaju verovatnoću svog opstanka i produžetka vrste.

Par 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.

Formalna definicija mašinskog učenja (Tom Mičel).

- Za jedan sistem VI kažemo da uči zadatu klasu zadataka T na osnovu iskustva E i zadate mere performansi P , ako se njegove performanse za rešavanje zadataka iz klase T , poboljšavaju sa iskustvom E .
- Da bi jedan problem učenja bio dobro definisan neophodno je identifikovati tri komponente:
 - klasu zadataka (T),
 - meru performanse koju treba poboljšati (P) i
 - izvor iskustva (E).
- Primer – problem učenja prepoznavanja rukom pisanih znakova.
 - Zadatak T je prepoznavanje i klasifikacija rukom pisanih reči u zadatim slikama
 - Mera performansi P je procenat tačno klasifikovanih reči.
 - Izvor iskustva E je baza rukom pisanih reči sa tačnom klasifikacijom (obučavajući skup).

Dominantni pristupi.

- U savremenoj oblasti mašinskog učenja dominiraju sledeći pristupi:
 - induktivno učenje,
 - analitičko učenje (analogija sa logikom),
 - učenje na slučajevima (*case-based learning* – analogija sa ljudskim pamćenjem),
 - neuralne mreže (analogija sa neurobiologijom),
 - genetski algoritmi (analogija sa evolucijom) i
 - hibridni modeli (kombinacija više pristupa).
- Od najvećeg značaja za tekuću praksu u domenu računarstva i VI je induktivno učenje.
- Suština ovog tipa učenja je učenje na osnovu raspoloživih primera.
 - U svakodnevnom jeziku to možemo da nazovemo: učenje iz iskustva (drugih).
- S obzirom na objekat učenja najopštije je učenje funkcionalnih, tj. ulazno-izlaznih preslikavanja.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Ključni elementi.

- Ključni elementi u induktivnom učenju funkcionalnih preslikavanja su:
 - nepoznato preslikavanje,
 - obučavajući skup ulazno-izlaznih parova i
 - skup hipoteza unutar koga biramo finalnu hipotezu putem algoritma obučavanja.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Ulazni podaci.

- Ulazni podaci su po pravilu n -dimenzioni vektori, koji se u literaturi pominju pod nazivima: vektori obeležja, uzorci, primeri i instance.
- Komponente ulaznih vektora se nazivaju obeležja ili atributi i mogu biti:
 - kontinualni (beskonačan broj vrednosti) ili
 - diskretni (konačan broj vrednosti).
- Ove komponente mogu biti tri različite prirode:
 - realni brojevi, npr. $x_i = 0,34$,
 - diskretni brojevi, npr. $x_i \in \{0, 1, 2, 3\}$ ili
 - kategorijalne varijable, npr. $x_i \in \{\text{crven, plav, zelen}\}$.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Izlazni prostor.

- Izlazni prostor može biti:
 - kategorijalan sa K distikntnih vrednosti, kada obučeni sistem obavlja klasifikaciju (izlaz se u ovom slučaju naziva oznaka, klasa, kategorija ili odluka),
 - realan, kada obučeni sistem realizuje regresiju (funkcionalni estimator) ili
 - vektorski, sa komponentama koje mogu biti realne ili kategorijalne varijable.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Obučavanje.

- Obučavanje se vrši na osnovu obučavajućeg skupa (N ulazno-izlaznih parova).
- Obučavanje možemo posmatrati kao pretragu u prostoru hipoteza u cilju izbora najpogodnije hipoteze koja je u saglasnosti sa obučavajućim skupom.
- Postoji više načina na koji se obučavajući skup može koristiti:
 - batch metod – celokupan obučavajući skup se koristi u izračunavanju hipoteza,
 - inkrementalni metod – koristi se sekvencijalno, po jedna instanca obučavajućeg skupa, koja modifikuje tekuću hipotezu i
 - on-line metod – korišćenje jedne instance obučavajućeg skupa u trenutku kada ona postane dostupna.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Vrste obučavanja.

- Obučavanje sa učiteljem, nadgledano obučavanje (*supervised learning*).
- Obučavanje bez učitelja, nenadgledano obučavanje (*unsupervised learning*).
- Obučavanje sa podsticanjem (*reinforcement learning*).

Induktivno empirijsko učenje funkcionalnih preslikavanja

Obučavanje sa učiteljem.

- Obučavajući skup je oblika $\{x_i, f(x_i)\}$, gde je $f(x_i)$ ciljna vrednost vektora obeležja x_i .
- Zadatak obučavanja je nalaženje aproksimacije za f .
- Mera performansi je kvalitet aproksimacije u tačkama koje ne pripadaju obučavajućem skupu.
- Obučavanje sa učiteljem može biti transduktivno i induktivno.
 - Transduktivno.
 - Test primeri bez ciljne klasifikacije se znaju već u trenutku obučavanja.
 - Cilj je da se sistem dobro ponaša samo na tim test primercima.
 - Induktivno.
 - Cilj je da se sistem dobro ponaša na svim primerima dobijenim iz iste raspodele kao i obučavajući skup.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Obučavanje bez učitelja.

- U obučavajućem skupu su prisutne samo ulazne instance x_i .
- Tipičan pristup je klasterovanje, odnosno grupisanje raspoloživih podataka u manji broj grupa unutar kojih su podaci sličniji u poređenju sa podacima iz ostalih grupa.
- Imenovanjem klastera se dolazi posrednim putem do označenih uzoraka.

Induktivno empirijsko učenje funkcionalnih preslikavanja

Obučavanje sa podsticanjem.

- Poreklo vodi iz teorije upravljanja u kome je dinamičko okruženje opisano trojkom (stanje, akcija, nagrada).
- U ovom tipu učenja potrebno je naučiti kako vršiti preslikavanja situacija u akcije a da se pri tome maksimizira nagrada.
- Za razliku od obučavanja sa učiteljem algoritmu obučavanja nije rečeno koje akcije da preduzima u datoj situaciji.
- Dobar primer da se shvati ovaj scenario je učenje igranja šaha.
 - Stanja su pozicije na tabli.
 - Akcije su mogući potezi za datu poziciju.
 - Nagrada za izabrani potez je pobeda.
 - Kazna je gubitak igre.
- Nagrada i kazna kasne u odnosu na trenutak izbora akcije, što je tipično za ovu vrstu obučavanja.

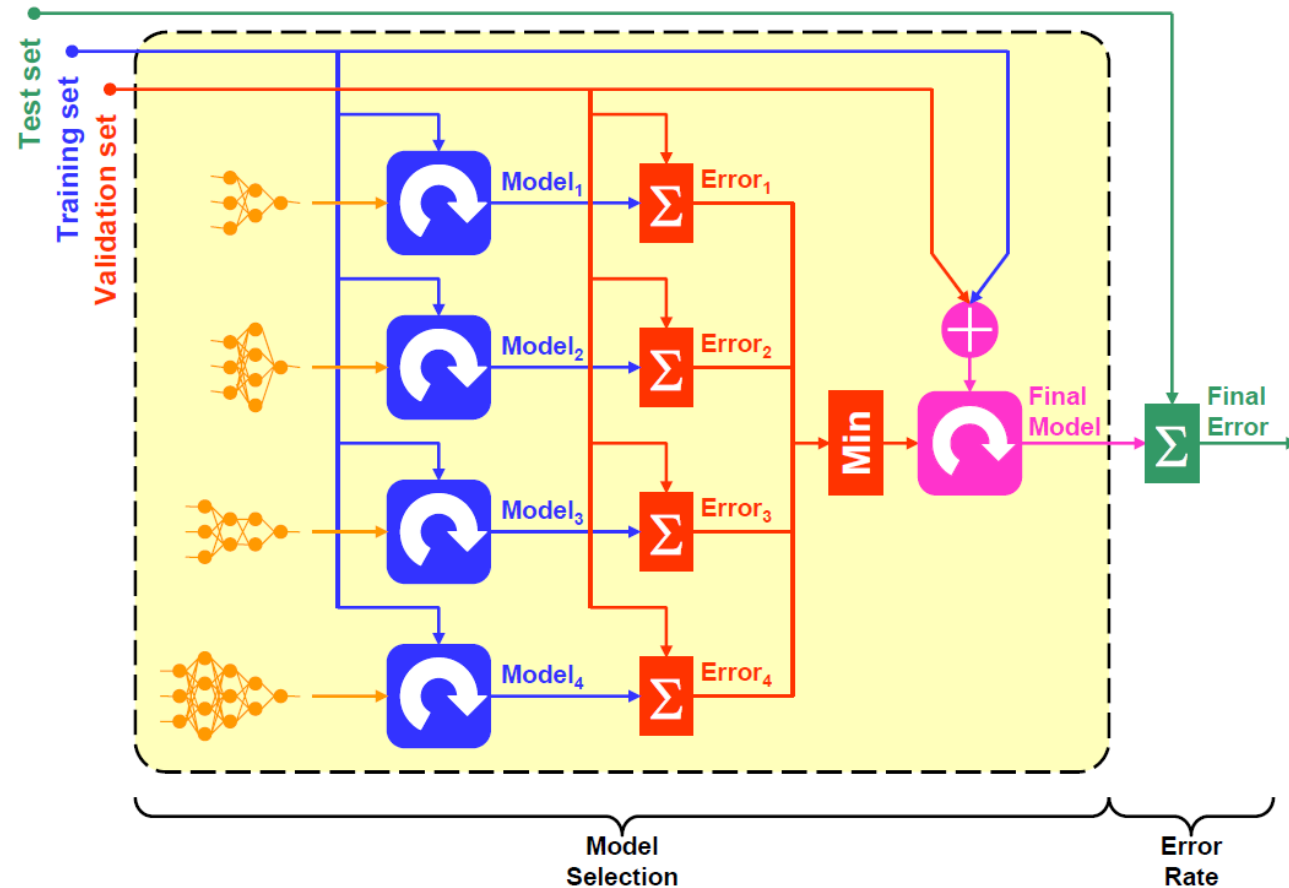
Generalizacija, podela na tri skupa i problem prenaučivosti

Obučavajući i test skup, generalizacija.

- Pitanje merenja kvaliteta obučenog sistema je jedno od ključnih pitanja mašinskog učenja.
- Izabrana hipoteza koja je najbolja moguća u odnosu na dati obučavajući skup:
 - može se pokazati kao dobra na novim primerima koji nisu učestvovali u obučavanju,
 - ali se može pokazati i kao loša.
- Opšteprihvaćena mera kvaliteta jednog sistema mašinskog učenja je njegova sposobnost generalizacije (ponašanja sistema na instancama koje nisu vidjene u fazi obučavanja).
- Stoga tipično sintezu prate dva skupa: obučavajući skup (trening skup) i test skup.
 - Test skup se koristi samo za procenu generalizacionih sposobnosti sintetisanog sistema.
 - Test skup se NE SME koristiti u fazi obučavanja.
- Uobičajene numeričke mere performansi su ili srednjekvadratna greška između izlaza sistema i ciljnih vrednosti ili ukupan broj grešaka.

Generalizacija, podela na tri skupa i problem prenaučnosti

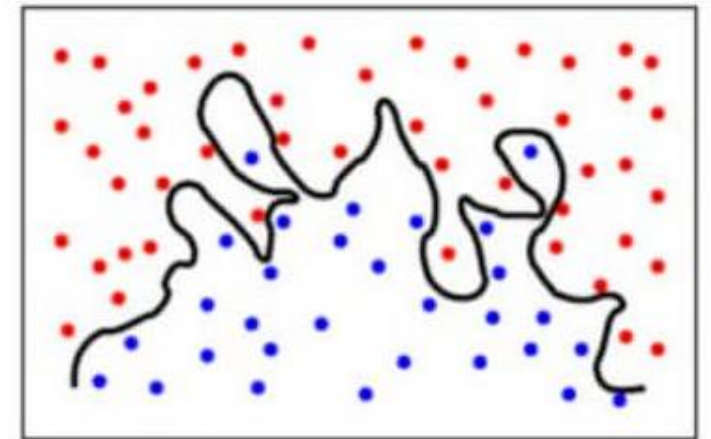
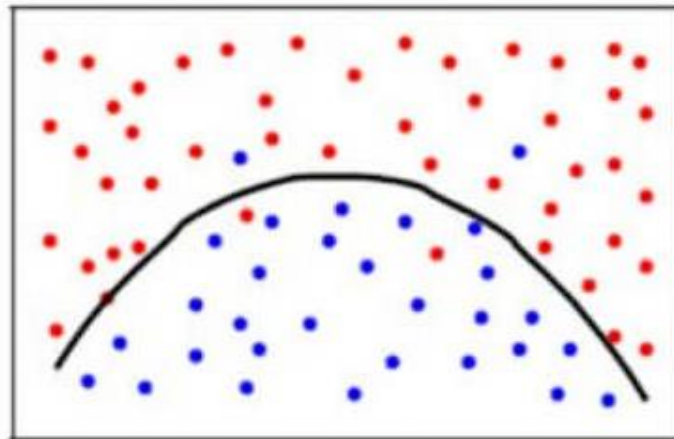
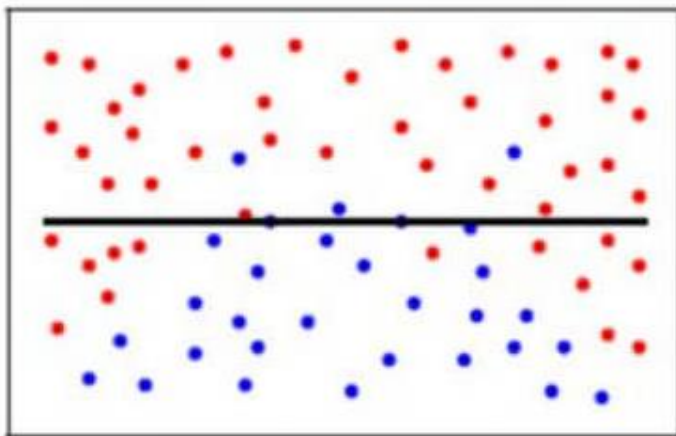
Podela na tri skupa.



Generalizacija, podela na tri skupa i problem prenaučivosti

Prenaučivost.

- Prilikom obuke je moguće da se model previše prilagodi specifičnostima obučavajućeg skupa.
 - U tom slučaju model daje loše rezultate kada se primeni na test skupu ili drugim podacima.
 - Ova pojava se naziva prenaučivost (engl. *overfitting*).
- Osnovni razlog zbog kog dolazi do prenaučivosti je bogatstvo prostora hipoteza, odnosno skupa dopustivih modela, u kome lakše naći model koji dobro odgovara podacima.



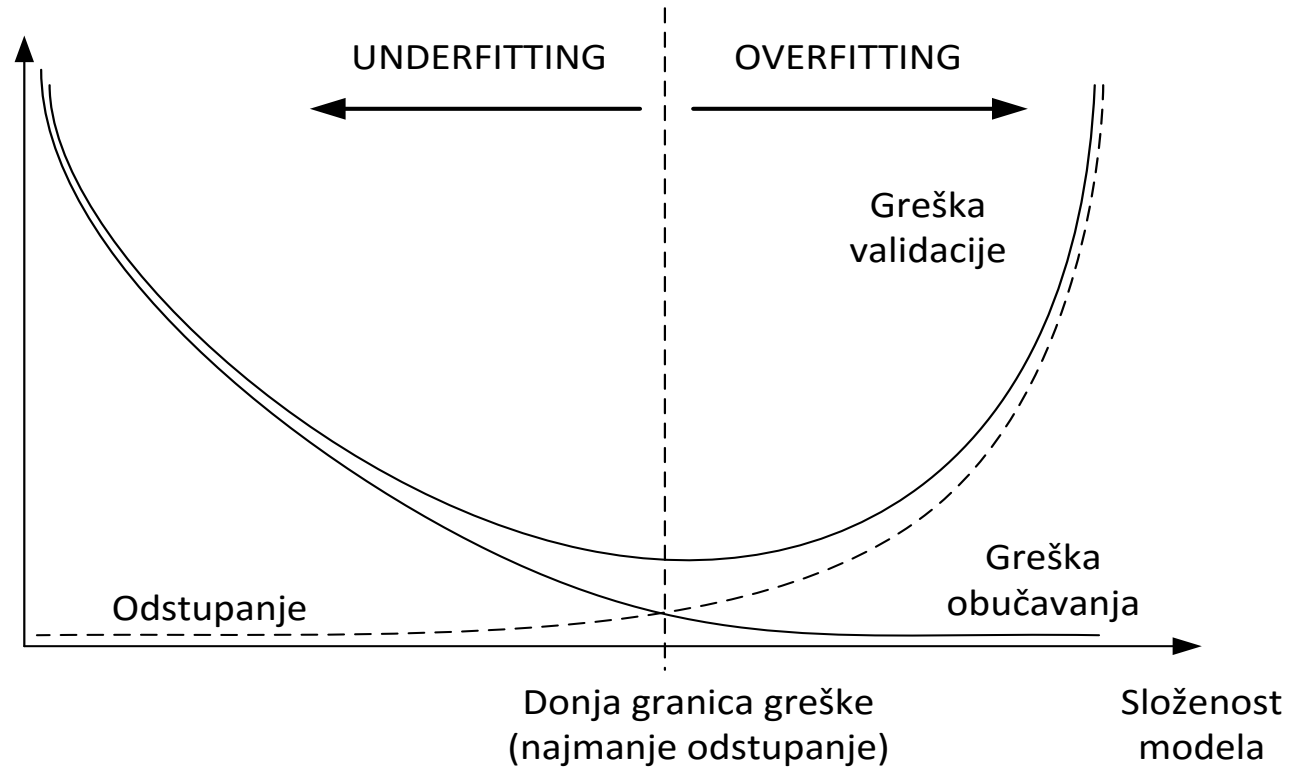
Generalizacija, podela na tri skupa i problem prenaučnosti

Prenaučnost.

- Primer *underfitting*-a.
 - Prilikom učenja dopuštaju se samo stabla odlučivanja dubine 1 koja testiraju samo jedan atribut svake instance.
 - Takva stabla ne mogu lako postići visoku preciznost klasifikacije.
- Primer *overfitting*-a.
 - Prilikom učenja dozvoljava se upotreba stabala odlučivanja proizvoljne dubine.
 - Moguće je naći stablo koje precizno opisuju svaku (čak i najnebitniju) specifičnost obučavajućih podataka.
 - Takvo stablo je u stanju da razlikuje bilo koje dve instance u obučavajućem skupu.
 - Takvo stablo postiže savršenu preciznost na obučavajućem skupu.
 - Ovakva stabla se u praksi pokazuju nepouzdanim!
 - Razlog: širi skupovi podataka ne moraju uvek imati sve specifičnosti obučavajućeg skupa!

Generalizacija, podela na tri skupa i problem prenaučnosti

Prenaučnost.



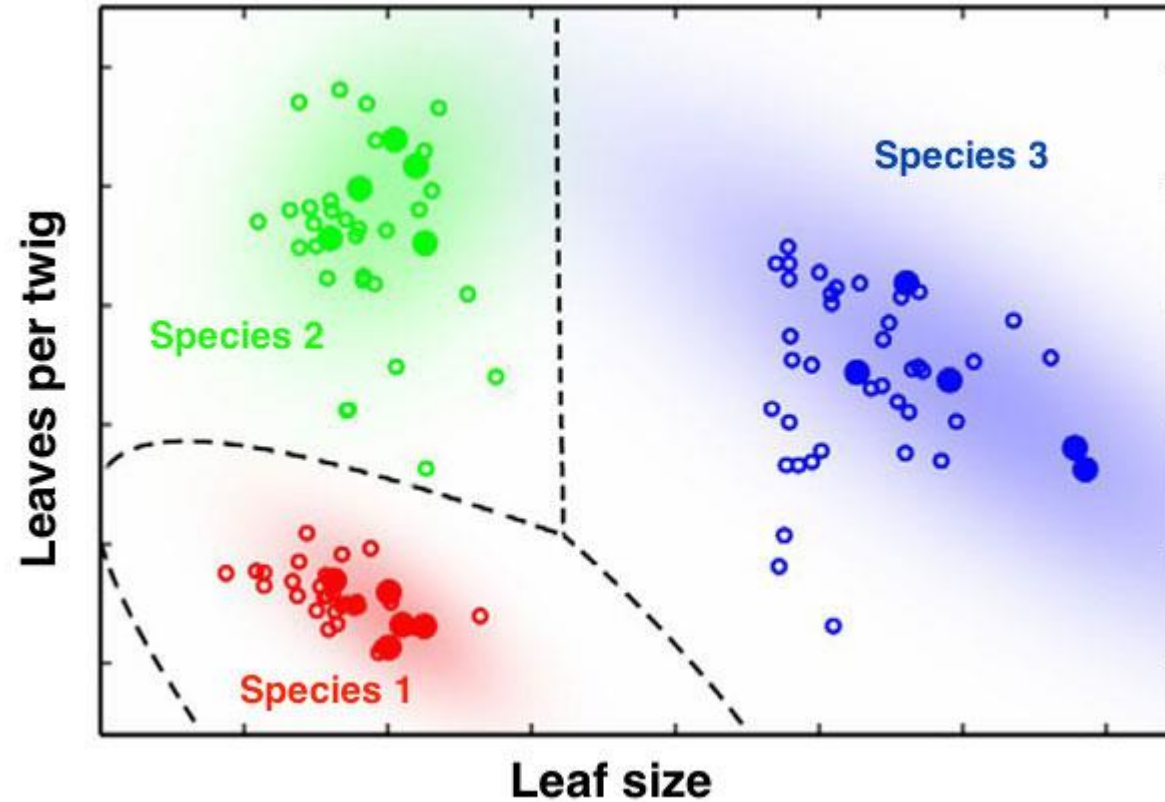
Šta je klasifikacija?

- Klasifikacija je razvrstavanje nepoznate instance u jednu od unapred ponuđenih klasa.
 - Svaka instanca može se predstaviti skupom njenih obeležja.
 - Takođe, svakoj instanci se može dodati kao obeležje i oznaka klase kojoj instanca pripada.

sepalength	sepalwidth	petallength	petalwidth	class
5.3	3.7	1.5	0.2	Iris-setosa
5	3.3	1.4	0.2	Iris-setosa
7	3.2	4.7	1.4	Iris-versicolor
6.2	2.8	4.8	1.8	Iris-virginica

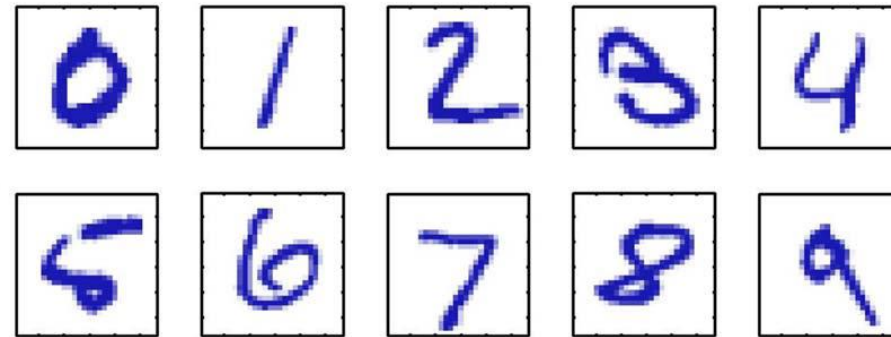
- Klasifikacija je određivanje vrednosti obeležja klase na osnovu preostalih obeležja instance.
- Postoji veliki broj metoda kojima se ovaj problem rešava: VNM, SVM, stabla odlučivanja, ...

Primer.



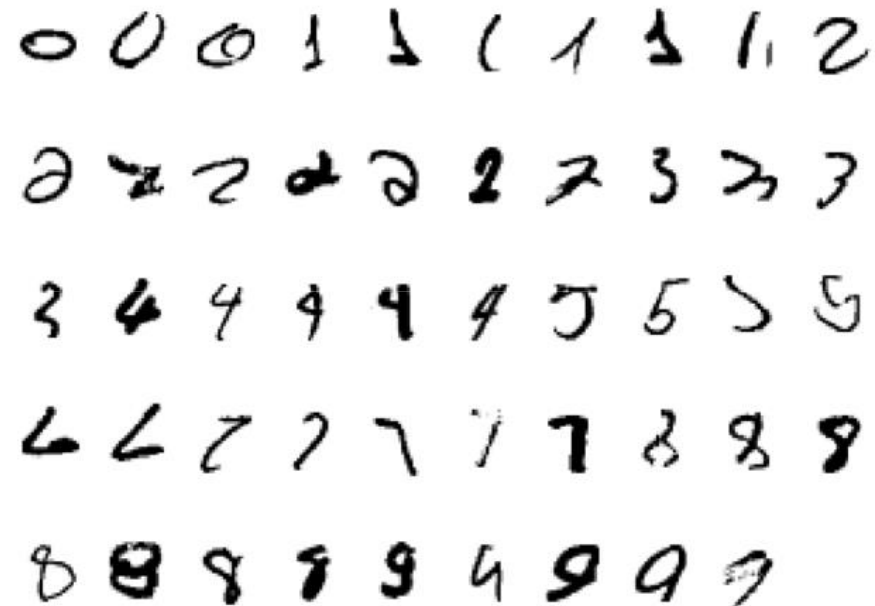
* Slika preuzeta sa Web lokacije: <https://astrobites.org/2015/04/15/c-3po-phd-machine-learning-in-astronomy/>

Primer klasifikacije – prepoznavanje rukopisa.



- Slike su pikseli dimenzije 28 x 28 piksela.
- Ulazna slika se predstavlja kao vektor: $x \in \mathbb{R}^{784}$
- Obučiti klasifikator $f(x)$ tako da $f: x \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
- Šta dalje?

Primer klasifikacije – prepoznavanje rukopisa.



- Rešavati kao problem nadgledanog učenja!
- Početi sa obučavajućim skupom, na primer 6000 instanci za svaku cifru.
- Može se postići visoka tačnost na test skupu (greška na testu oko 0.4%).

- Beleške pripremljene prema knjizi – Milan Milosavljević: „Veštačka inteligencija“. Univerzitet Singidunum, Beograd.

Pitanja su dobrodošla.