

Ako predpovedať iné veci ako počasie

Náhodný les

Peter Novotný

FMFI UK
Bratislava

Akadémia Trojstenu

Obsah

- 1 Strojové učenie
- 2 Rozhodovacie stromy
- 3 Súťaže v strojovom učení
- 4 Záver

Úvod

- Posledné desaťročia je svet zamorovaný množstvom dát.

Úvod

- Posledné desaťročia je svet zamorovaný množstvom dát.
- Vznikla úloha orientovať sa v dátach a vydolovať z nich užitočné údaje.

Úvod

- Posledné desaťročia je svet zamorovaný množstvom dát.
- Vznikla úloha orientovať sa v dátach a vydolovať z nich užitočné údaje.
- Spolu s dátami našťastie vznikli aj počítače, ktoré dokážu s obrovským množstvom dát pracovať.

Úvod

- Posledné desaťročia je svet zamorovaný množstvom dát.
- Vznikla úloha orientovať sa v dátach a vydolovať z nich užitočné údaje.
- Spolu s dátami našťastie vznikli aj počítače, ktoré dokážu s obrovským množstvom dát pracovať.
- Ešte predtým našťastie vznikla matematika a v rámci nej štatistika.

Príklady problémov

Čo majú spoločné nasledovné otázky?

Príklady problémov

Čo majú spoločné nasledovné otázky?

- Zaberie pacientovi liek?

Príklady problémov

Čo majú spoločné nasledovné otázky?

- Zaberie pacientovi liek?
- Bude klient schopný splácať pôžičku?

Príklady problémov

Čo majú spoločné nasledovné otázky?

- Zaberie pacientovi liek?
- Bude klient schopný splácať pôžičku?
- Je e-mail, ktorý práve prišiel do schránky, spam?

Príklady problémov

Čo majú spoločné nasledovné otázky?

- Zaberie pacientovi liek?
- Bude klient schopný splácať pôžičku?
- Je e-mail, ktorý práve prišiel do schránky, spam?
- Je na obrázku, ktorý zaznamenal teleskop, asteroid?

Príklady problémov

Čo majú spoločné nasledovné otázky?

- Zaberie pacientovi liek?
- Bude klient schopný splácať pôžičku?
- Je e-mail, ktorý práve prišiel do schránky, spam?
- Je na obrázku, ktorý zaznamenal teleskop, asteroid?
- V akom jazyku rozpráva ten človek v telefóne?

Príklady problémov

Čo majú spoločné nasledovné otázky?

- Zaberie pacientovi liek?
- Bude klient schopný splácať pôžičku?
- Je e-mail, ktorý práve prišiel do schránky, spam?
- Je na obrázku, ktorý zaznamenal teleskop, asteroid?
- V akom jazyku rozpráva ten človek v telefóne?
- ...

Strojové učenie

Na rovnaké otázky (len pre iné „subjekty“) máme množstvo správnych odpovedí v databáze z minulosti (tzv. tréningovú množinu).

Strojové učenie

Na rovnaké otázky (len pre iné „subjekty“) máme množstvo správnych odpovedí v databáze z minulosti (tzv. trénovaciu množinu).

Ak v trénovacej množine nájdeme subjekty s podobnými vlastnosťami, dá sa očakávať, že náš subjekt sa bude správať podobne.

Strojové učenie

Na rovnaké otázky (len pre iné „subjekty“) máme množstvo správnych odpovedí v databáze z minulosti (tzv. trénovaciu množinu).

Ak v trénovacej množine nájdeme subjekty s podobnými vlastnosťami, dá sa očakávať, že náš subjekt sa bude správať podobne.

Lenže čo znamená „podobné vlastnosti“ v tom nesmiernom mori rozličných údajov? Akým spôsobom zvoliť odpoveď pre náš subjekt?

Strojové učenie

Na rovnaké otázky (len pre iné „subjekty“) máme množstvo správnych odpovedí v databáze z minulosti (tzv. trénovaciu množinu).

Ak v trénovacej množine nájdeme subjekty s podobnými vlastnosťami, dá sa očakávať, že náš subjekt sa bude správať podobne.

Lenže čo znamená „podobné vlastnosti“ v tom nesmiernom mori rozličných údajov? Akým spôsobom zvoliť odpoveď pre náš subjekt?

Týmto sa zaoberá *strojové učenie*.

Rôzne metódy strojového učenia

Strojové učenie je v súčasnosti mohutne expandujúca a rozvíjajúca sa disciplína. Zahŕňa množstvo rôznych prístupov, napr.:

Rôzne metódy strojového učenia

Strojové učenie je v súčasnosti mohutne expandujúca a rozvíjajúca sa disciplína. Zahŕňa množstvo rôznych prístupov, napr.:

- rôzne naivné postupy,

Rôzne metódy strojového učenia

Strojové učenie je v súčasnosti mohutne expandujúca a rozvíjajúca sa disciplína. Zahŕňa množstvo rôznych prístupov, napr.:

- rôzne naivné postupy,
- rozhodovacie stromy,

Rôzne metódy strojového učenia

Strojové učenie je v súčasnosti mohutne expandujúca a rozvíjajúca sa disciplína. Zahŕňa množstvo rôznych prístupov, napr.:

- rôzne naivné postupy,
- rozhodovacie stromy,
- neurónové siete,

Rôzne metódy strojového učenia

Strojové učenie je v súčasnosti mohutne expandujúca a rozvíjajúca sa disciplína. Zahŕňa množstvo rôznych prístupov, napr.:

- rôzne naivné postupy,
- rozhodovacie stromy,
- neurónové siete,
- rôzne typy regresíí,

Rôzne metódy strojového učenia

Strojové učenie je v súčasnosti mohutne expandujúca a rozvíjajúca sa disciplína. Zahŕňa množstvo rôznych prístupov, napr.:

- rôzne naivné postupy,
- rozhodovacie stromy,
- neurónové siete,
- rôzne typy regresíí,
- ...

Vytvorenie trénovacej množiny

Ukážeme si na konkrétnom príklade jednu z metód strojového učenia, ktorá sa veľmi jednoducho implementuje.

Vytvorenie trénovacej množiny

Ukážeme si na konkrétnom príklade jednu z metód strojového učenia, ktorá sa veľmi jednoducho implementuje.
Budeme pracovať s reálnymi dátami, ktoré sa týkajú všetkých účastíkov tu v posluchárni.

Vytvorenie trérovacej množiny

Ukážeme si na konkrétnom príklade jednu z metód strojového učenia, ktorá sa veľmi jednoducho implementuje.

Budeme pracovať s reálnymi dátami, ktoré sa týkajú všetkých účastíkov tu v posluchárni.

V praxi je dôležité dbať na to, aby trérovacia množina bola „dobrá“.

Vytvorenie trénovacej množiny

Ukážeme si na konkrétnom príklade jednu z metód strojového učenia, ktorá sa veľmi jednoducho implementuje.

Budeme pracovať s reálnymi dátami, ktoré sa týkajú všetkých účastíkov tu v posluchárni.

V praxi je dôležité dbať na to, aby trénovacia množina bola „dobrá“.

`http://skmo.sk/akademia`

Pestovanie stromu

Našou úlohou je trénovaciu množinu rozdeliť (postupným rozkonárovaním stromu) na časti (listy), v rámci ktorých už budú mať všetky subjekty rovnakú „odpoveď“.

Pestovanie stromu

Našou úlohou je trénovaciu množinu rozdeliť (postupným rozkonárovaním stromu) na časti (listy), v rámci ktorých už budú mať všetky subjekty rovnakú „odpoveď“.

V každom kroku sa snažíme rozkonáriť aktuálnu množinu subjektov „čo najlepšie“.

Pestovanie stromu

Našou úlohou je trénovaciu množinu rozdeliť (postupným rozkonárovaním stromu) na časti (listy), v rámci ktorých už budú mať všetky subjekty rovnakú „odpoveď“.

V každom kroku sa snažíme rozkonáriť aktuálnu množinu subjektov „čo najlepšie“.

Keď však vytvoríme iba jeden strom, veľmi často je „pretrénovaný“.

Náhodný les

V praxi oveľa lepšie výsledky dostávame, keď vytvoríme množstvo stromov.

Náhodný les

V praxi oveľa lepšie výsledky dostávame, keď vytvoríme množstvo stromov.

- Každý strom vyrastie na mierne odlišnej množine subjektov.

Náhodný les

V praxi oveľa lepšie výsledky dostávame, keď vytvoríme množstvo stromov.

- Každý strom vyrastie na mierne odlišnej množine subjektov.
- V každom kroku do rozkonárovania vložíme náhodný prvok.

Náhodný les

V praxi oveľa lepšie výsledky dostávame, keď vytvoríme množstvo stromov.

- Každý strom vyrastie na mierne odlišnej množine subjektov.
- V každom kroku do rozkonárovania vložíme náhodný prvok.
- Na konci jednoducho spriemerujeme výsledky z jednotlivých stromov.

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží

Kaggle: <https://www.kaggle.com/>

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží
- 2 – 3 týždne jedna súťaž

Kaggle: <https://www.kaggle.com/>

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží
- 2 – 3 týždne jedna súťaž
- nutnosť odovzdať funkčný program

Kaggle: <https://www.kaggle.com/>

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží
- 2 – 3 týždne jedna súťaž
- nutnosť odovzdať funkčný program

Kaggle: <https://www.kaggle.com/>

- v každom momente otvorených niekoľko súťaží

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží
- 2 – 3 týždne jedna súťaž
- nutnosť odovzdať funkčný program

Kaggle: <https://www.kaggle.com/>

- v každom momente otvorených niekoľko súťaží
- jedna súťaž trvá často aj viac mesiacov

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží
- 2 – 3 týždne jedna súťaž
- nutnosť odovzdať funkčný program

Kaggle: <https://www.kaggle.com/>

- v každom momente otvorených niekoľko súťaží
- jedna súťaž trvá často aj viac mesiacov
- možnosť súťažiť v tímoch

TopCoder: <https://community.topcoder.com/longcontest/>

- za rok 5 až 10 súťaží
- 2 – 3 týždne jedna súťaž
- nutnosť odovzdať funkčný program

Kaggle: <https://www.kaggle.com/>

- v každom momente otvorených niekoľko súťaží
- jedna súťaž trvá často aj viac mesiacov
- možnosť súťažiť v tímoch
- väčšia konkurencia

