Masarykova univerzita
Fakulta informatiky

Predvídánie výkyvov v rezerváciach leteniek

Diplomová práca

Robert Kuska

Brno, jar 2016

Masarykova univerzita
Fakulta informatiky

Predvídánie výkyvov v rezerváciach leteniek

Diplomová práca

Robert Kuska

Brno, jar 2016

Namiesto tejto stránky vložte kópiu oficiálneho podpísaného zadania práce a prehlásenie autora školského diela.

Prehlásenie

Prehlasujem, že táto diplomová práca je mojím pôvodným autorským dielom, ktoré som vypracoval samostatne. Všetky zdroje, pramene a literatúru, ktoré som pri vypracovaní používal alebo z nich čerpal, v práci riadne citujem s uvedením úplného odkazu na príslušný zdroj.

Robert Kuska

Vedúci práce: doc. RNDr. Aleš Horák Ph.D

Poďakovanie

Chcel by som poďakovať svojej priateľke, ktorá ma presvedčila nech nevzdávam svoj nekonečný boj s písaním diplomovej práce.

Zhrnutie

Hlavným cieľom tejto diplomovej práce je praktické využitie regresných modeloov a ich aplikácia na predvídanie výkyvov v skutočných rezerváciach leteniek v spojení s informáciami o událostiach. Súčasťou práce je podrobný prehľad a vysvetlenie najpoužívanejších algoritmov v oblasti strojového učenia pre úlohy učenia s učiteľom.

V praktickej časti práce sú použité vybrané algoritmy na analýzu reálnych dát. Ako vstup su použíté dáta z metavyhľadávačov leteniek pre 20 najvyťaženejších leteckých tras s cieľovým alebo zdrojovým letiskom v Londýne. Aplikácia detekuje a predvída mimoriadne výkyvy v počte rezervácií v súvislosti s externými zdrojmi údajov o akciach, ktoré slúžia pre nájdenie korelacií a predikcií.

Kľúčové slová

lineárna regresia, machine learning, Londýn, lety

Obsah

1 Úvod **1**

2 Použitý softvér **3**
2.1 *Scikit-learn* 3
2.2 *Pandas* 3
2.3 *Seaborn* 3

3 Lineárna regresia **5**
3.1 *Metóda najmenších štvorcov* 5
3.2 *Gradient Descent* 5
3.3 *Bayesova Regrese* 5
3.4 *Lasso* 5

4 Floats and references **7**

5 Inserting the bibliography **9**

Literatúra **11**

6 Inserting the index **13**

Register **15**

A An appendix **17**

Zoznam tabuliek

Zoznam obrázkov

1 Úvod

Žijeme v dobe kedy je letecká doprava vďaka rozšíreniu nízkonákladových letov viac dostupná pre bežných ľudí zo strednej triedy. Ľudia využívajú leteckú dopravu viac než kedykoľvek predtým, z dôvodu rýchlosti prepravy, ktorej iný spôsob dopravy momentálne nemôže konkurovať a cena tohoto spôsobu prepravy ôsob je taktiež konkurencie schopná tradičným druhom cestovania. S rozmachom leteckej prepravy taktiež prichádzajú ďalšie výzvy, ktoré súvisia s lepšovaním komfortnosti a spoľahlivosti. Jednou z mnohých výziev je napríklad analýza a predikcia vyťaženosti leteckých liniek z dôvodu prispôsobenia ponuky letov pre čo najefektívnejšie pokrytie dopytu trhu. Táto práca sa bude snažiť pomocou regresných modelov predpovedať dopyt zákazníkov pre lety z alebo do Londýna a taktiež sa bude snažiť predikovať výkyvy vo vyhľadávaní, ktoré nezapadajú do predikovaných štatistických modelov. Tieto výkyvy sa budem snažiť priradiť k externým událostiam, ako sú napríklad rôzne športové události alebo konferencie.

2 Použitý softvér

V tejto kapitole si priblížimé softvér a frameworky, ktoré boli použité pri výpočte modelov a analýzu dát. Celá diplomová práca bola napísaná za pomoci programovacieho jazyka Python. Jazyk Python bol použitý z dôvodu veľkého zastúpenia rozcličných frameworkov určených na použitie v strojovom učení a taktiež v dátovej analýze.

Z frameworkov boli použité nasledujúce:

- scikit-learn - Scikit-learn je framework zameraný na strojové učenie, obsahuje veľké množstvo naimplementovaných algoritmov používaných v oblasti strojového učenia
- pandas - Framework Pandas slúži na lepšiu prácu s dátami, je inšpirovaný DataFrame štruktúrou z jazyka R
- seaborn - Na vytváranie grafov bol použitý framework Seaborn

2.1 Scikit-learn

2.2 Pandas

2.3 Seaborn

3 Lineárna regresia

Lineárna regresie je model ktorý predpokladá lineárnu závislosť medzi vstupnými premennými a jednou výstupnou premennou. Výstupná premenná je teda spočítaná z linearných kombinácií vstupných premenných. V prípade, že máme iba jednu vstupnú premennú hovoríme o jednoduche lineárnej regresii, v prípade viacerých vstupných premenných sa jedná o viacnásobnú lineárnu regresiu. Na výpočet regresie sa používajú viaceré metódy, my si predstavíme tie, ktoré sa používajú najčastejšie a sú taktiež použité pri spracovaní našich dát.

3.1 Metóda najmenších štvorcov

V prípade, že máme viac než jeden vstupný parameter, môžme k vytvoreniu lineárneho modelu použit metódu najmenších štvrcov. Funguje tak, že k danej regresnej priamke, ktorá vedie cez dáta, spočítame vzdialenosť od každého bodu k regresnej priamke, túto vzdialenosť umocníme a spočítame dohromady s ostatnými umocnenými vzdialenosťami. Ako lineárna priamka nám môže poslúžiť obyčajná priamka, parabola alebo obecný polynom x stupňa.

3.2 Gradient Descent

3.3 Bayesova Regrese

3.4 Lasso

4 Floats and references

5 Inserting the bibliography

After loading the `biblatex` package and linking a bibliography database file to the document using the `\addbibresource` command, you can start citing the entries. This is just dummy text [1] lightly sprinkled with citations [2, p. 123]. Several sources can be cited at once [3–5]. „Fundamental Algorithms“ was written by Knuth in 1973. We can also produce Knuth [1] or (Knuth [1], 1973). The full bibliographic citation is: *Donald E. Knuth. „Fundamental Algorithms“. In: Second. Vol. 1. The Art of Computer Programming. This is a full INBOOK entry. Reading, Massachusetts: Addison-Wesley, Oct. 1973. Chap. 1.2, pp. 10–119*. We can easily insert a bibliographic citation into the footnote[1].

The `\nocite` command will not generate any output, but it will insert its argument into the bibliography. The `\nocite{*}` command will insert all the records in the bibliography database file into the bibliography. Try uncommenting the command and watch the bibliography section come apart at the seams.

When typesetting the document for the first time, citing a `work` will expand to [**work**] and the `\printbibliography` command will produce no output. It is now necessary to generate the bibliography by running `biber fi-pdflatex.bcf` from the command line and then by typesetting the document again twice. During the first run, the bibliography section and the citations will be typeset, and in the second run, the bibliography section will appear in the table of contents.

The `biber` command needs to be executed from within the directory, where the LaTeX source file is located. In Windows, the command line can be opened in a directory by holding down the Shift key and by clicking the right mouse button while hovering the cursor over a directory. Select the Open Command Window Here option in the context menu that opens shortly afterwards.

With online services – such as Overleaf – all commands are executed automatically.

---

1. Donald E. Knuth. „Fundamental Algorithms“. In: Second. Vol. 1. The Art of Computer Programming. This is a full INBOOK entry. Reading, Massachusetts: Addison-Wesley, Oct. 1973. Chap. 1.2, pp. 10–119.

Literatúra

[1] Donald E. Knuth. „Fundamental Algorithms“. In: Second. Vol. 1. The Art of Computer Programming. This is a full INBOOK entry. Reading, Massachusetts: Addison-Wesley, Oct. 1973. Chap. 1.2, pp. 10–119.

[2] Daniel D. Lincoll. „Semigroups of Recurrences“. In: *High Speed Computer and Algorithm Organization*. Ed. by David J. Lipcoll, D. H. Lawrie, and A. H. Sameh. Third. Fast Computers 23. This is a full INCOLLECTION entry. New York: Academic Press, Sept. 1977. Chap. 3, pp. 179–183.

[3] David J. Lipcoll, D. H. Lawrie, and A. H. Sameh, eds. *High Speed Computer and Algorithm Organization*. Third. Fast Computers 23. This is a cross-referenced BOOK (collection) entry. New York: Academic Press, Sept. 1977.

[4] *The Definitive Computer Manual*.

[5] Larry Manmaker. *The Definitive Computer Manual*. Silver. This is a full MANUAL entry. Chips-R-Us. Silicon Valley, Apr. 1986.

[6] Jill C. Knvth. *The Programming of Computer Art*. Vernier Art Center. This is a full BOOKLET entry. Stanford, California, Feb. 1988.

6 Inserting the index

After using the `\makeindex` macro and loading the `makeidx` package that provides additional indexing commands, index entries can be created by issuing the `\index` command. It is possible to create ranged index entries, which will encompass a span of text. To insert complex typographic material – such as α or TEX – into the index, you need to specify a text string, which will determine how the entry will be sorted. It is also possible to create hierarchal entries.

After typesetting the document, it is necessary to generate the index by running

```
texindy -I latex -C utf8 -L ⟨locale⟩ fi-pdflatex.idx
```

from the command line, where ⟨*locale*⟩ corresponds to the main locale of your thesis – such as `english`, and then typesetting the document again.

The `texindy` command needs to be executed from within the directory, where the LATEX source file is located. In Windows, the command line can be opened in a directory by holding down the Shift key and by clicking the right mouse button while hovering the cursor over a directory. Select the Open Command Window Here option in the context menu that opens shortly afterwards.

With online services – such as Overleaf – the commands are executed automatically, although the locale may be erroneously detected, or the `makeindex` tool (which is only able to sort entries that contain digits and letters of the English alphabet) may be used instead of `texindy`. In either case, the index will be ill-sorted.

Register

α, 13

dummy text, 13

TEX, 13

vehicles
    speed cars, 13
    trucks, 13

A An appendix

Here you can insert the appendices of your thesis.