

# Malá velká data

---

Aleš Horák, Jakub Šuraň, Michal Štefaňák, Filip Svoboda

# Obsah

---

- Cíle
- Teorie
- Kaggle
- Výsledky
- Závěr



# Obsah

---

- **Cíle**
- Teorie
- Kaggle
- Výsledky
- Závěr

# Cíle

---

- Seznámení s platformou kaggle
- Poznání základních metod strojového učení
- Vytvoření programu na zpracování dat o průhledných polovodičích

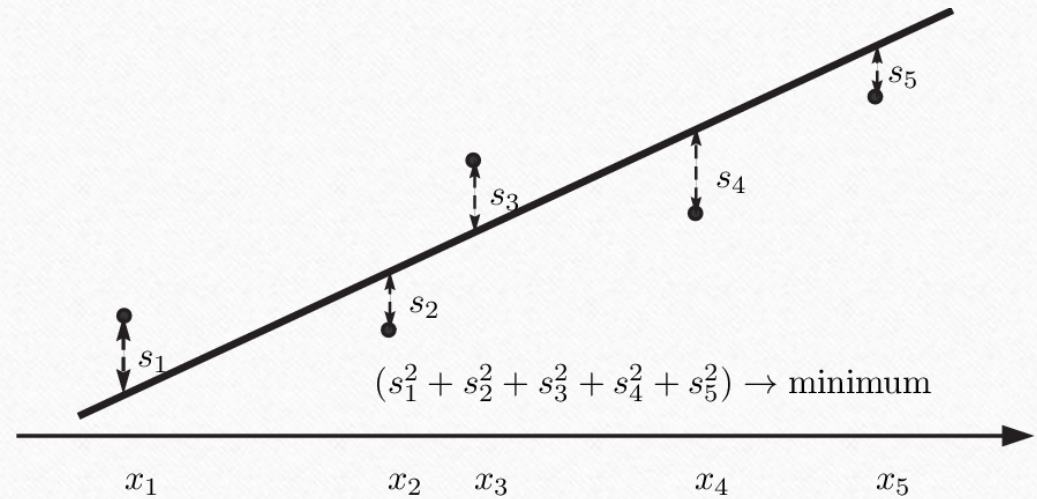
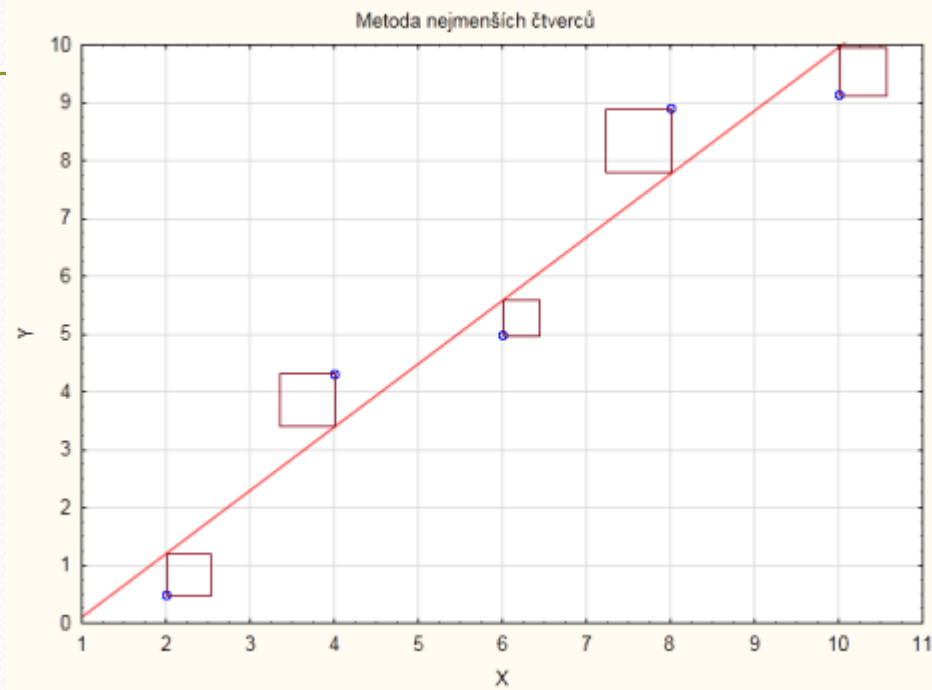
# Obsah

---

- Cíle
- **Teorie**
- Kaggle
- Výsledky
- Závěr



# Teorie



Metoda nejmenších čtverců

# Obsah

---

- Cíle
- Teorie
- **Kaggle**
- Výsledky
- Závěr

# Kaggle

- Platforma, zabývající se zpracování dat
- Patří pod google
- Soutěže mezi experty z celého světa



|                                                                                       |                                                                                                                                                                                                                   |                           |
|---------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|
|    | <b>Home Credit Default Risk</b><br>Can you predict how capable each applicant is of repaying a loan?<br><i>Featured</i> · 2 months to go · home, banking, tabular data                                            | \$70,000<br>2,898 teams   |
|    | <b>Santander Value Prediction Challenge</b><br>Predict the value of transactions for potential customers.<br><i>Featured</i> · 2 months to go · finance, banking                                                  | \$60,000<br>114 teams     |
|    | <b>TrackML Particle Tracking Challenge</b><br>High Energy Physics particle tracking in CERN detectors<br><i>Featured</i> · 2 months to go · physics, tabular data                                                 | \$25,000<br>434 teams     |
|    | <b>The 2nd YouTube-8M Video Understanding Challenge</b><br>Can you create a constrained-size model to predict video labels?<br><i>Featured</i> · 2 months to go · video data, object labeling                     | \$25,000<br>249 teams     |
|    | <b>Avito Demand Prediction Challenge</b><br>Predict demand for an online classified ad<br><i>Featured</i> · 8 days to go · tabular data, image data, text data                                                    | \$25,000<br>1,824 teams   |
|   | <b>House Prices: Advanced Regression Techniques</b><br>Predict sales prices and practice feature engineering, RFs, and gradient boosting<br><i>Getting Started</i> · Ongoing · tabular data, regression           | Knowledge<br>5,004 teams  |
|  | <b>Digit Recognizer</b><br>Learn computer vision fundamentals with the famous MNIST data<br><i>Getting Started</i> · Ongoing · tabular data, image data, multiclass classification, object identification         | Knowledge<br>2,701 teams  |
|  | <b>Titanic: Machine Learning from Disaster</b><br>Start here! Predict survival on the Titanic and get familiar with ML basics<br><i>Getting Started</i> · Ongoing · tutorial, tabular data, binary classification | Knowledge<br>11,346 teams |
|  | <b>ImageNet Object Detection Challenge</b><br>Identify and label everyday objects in images                                                                                                                       | Knowledge<br>0 teams      |



# Obsah

---

- Cíle
- Teorie
- Kaggle
- **Výsledky**
- Závěr

# Výsledek

|                                                                                                              |        |        |                          |
|--------------------------------------------------------------------------------------------------------------|--------|--------|--------------------------|
| <b>reply.csv</b><br>3 hours ago by <a href="#">Michal Štefaňák</a><br><a href="#">add submission details</a> | 0.0710 | 0.0558 | <input type="checkbox"/> |
| <b>reply.csv</b><br>4 hours ago by <a href="#">dzony</a><br><a href="#">add submission details</a>           | 0.0881 | 0.0785 | <input type="checkbox"/> |
| <b>Least_Squares.csv</b><br>6 hours ago by <a href="#">dzony</a><br><a href="#">add submission details</a>   | 0.1330 | 0.1205 | <input type="checkbox"/> |
| <b>sample_submission.csv</b><br>a day ago by <a href="#">dzony</a>                                           | 0.2219 | 0.2077 | <input type="checkbox"/> |

# Obsah

---

- Cíle
- Teorie
- Kaggle
- Výsledky
- **Závěr**



# Závěr

---

- Základ práce s Kaggle
- Seznámení s MATLAB
- Osvojení jednoduchých metod strojového učení

# Zdroje

---

- <https://www.kaggle.com/c/nomad2018-predict-transparent-conductors>
- <https://is.muni.cz/do/rect/el/estud/prif/ps15/statistika/web/pages/regres-anal.html>
- <http://user.mendelu.cz/marik/mat-web/mat-webse24.html>