

Predictive Modelling with Python

Jure Žabkar

Contents

Software installation

Introduction to scikit-learn

Artificial data sets, illustration of basic regression and classification techniques



Regression & Classification

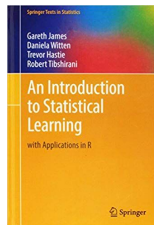
Individual work: data preparation, Visualization, Modelling, Feature selection, Evaluation

Installation

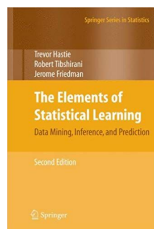


[Install conda](#)

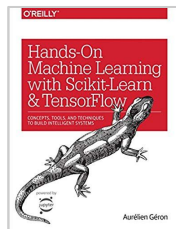
Readings



James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). ***An introduction to statistical learning*** (Vol. 6). New York: Springer.



Friedman, J., Hastie, T., & Tibshirani, R. (2009). ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction***. Springer Series in Statistics.



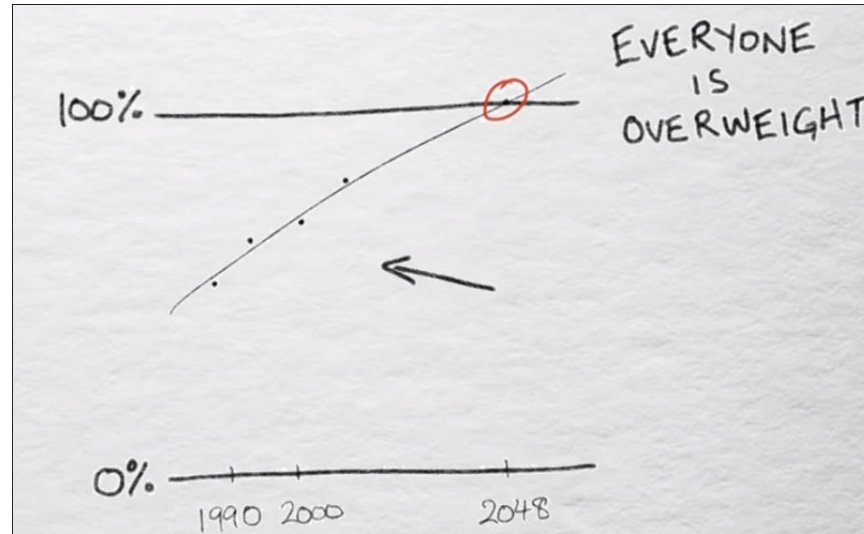
Geron, A. (2017). ***Hands-on machine learning with Scikit-Learn and TensorFlow***. O'Reilly. (there's also the 2nd edition of this book)

What will you learn?

- How to import the data
- Data preprocessing & visualization
- Computing basic data set statistics
- Basic regression and classification with sklearn
- How to tune the parameters of ML algorithms
- Proper evaluation

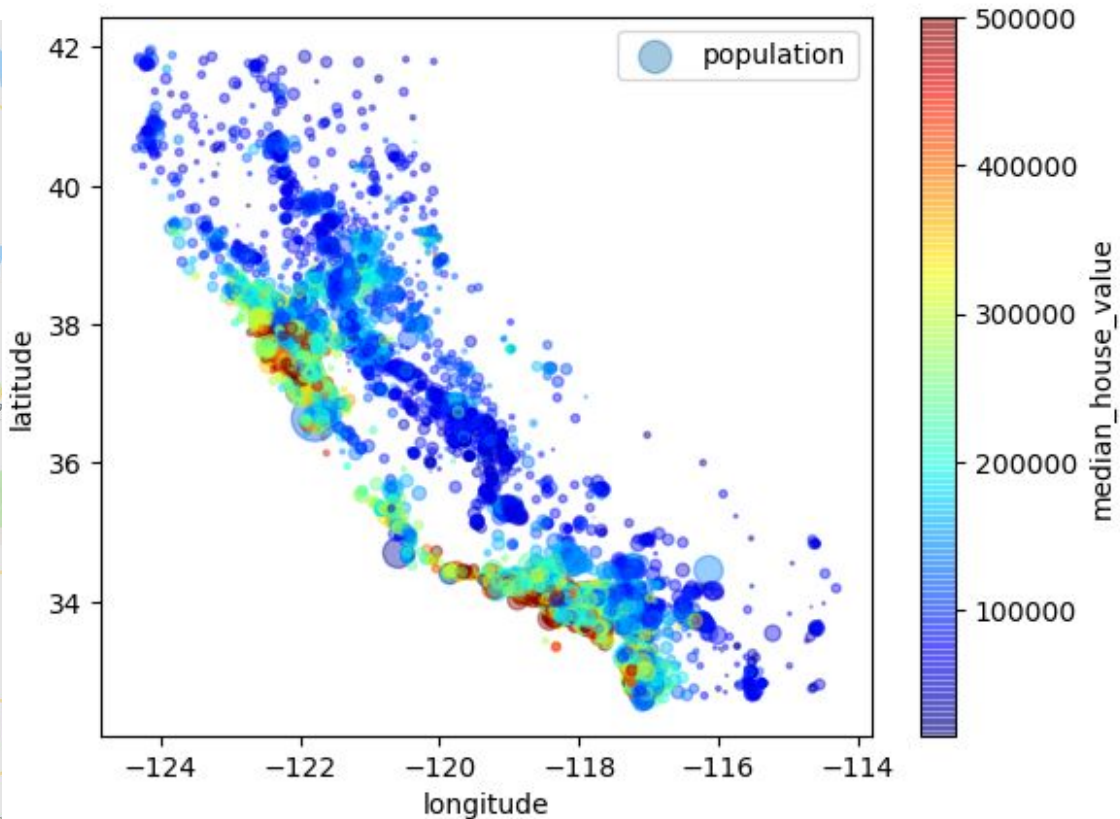
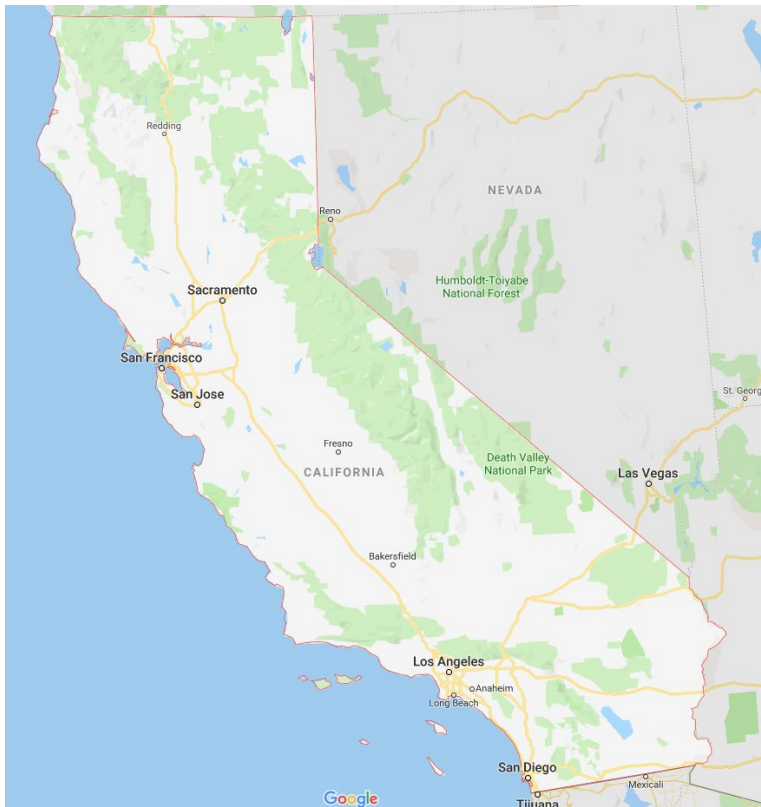
Obesity apocalypse

abcNEWS: "By 2048, all American adults would become overweight or obese."

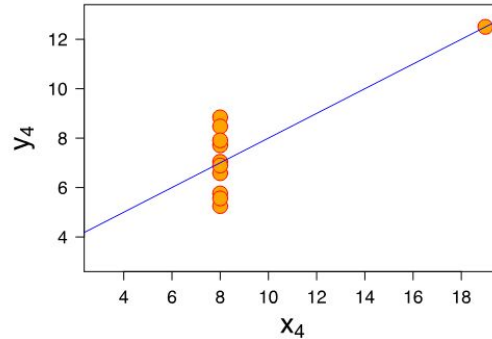
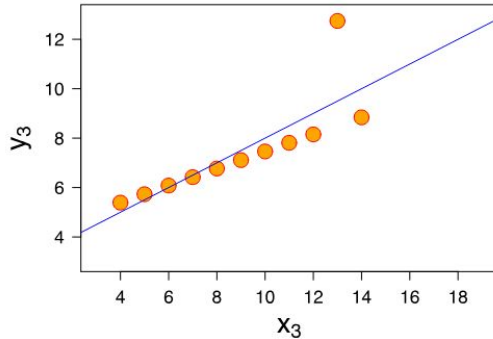
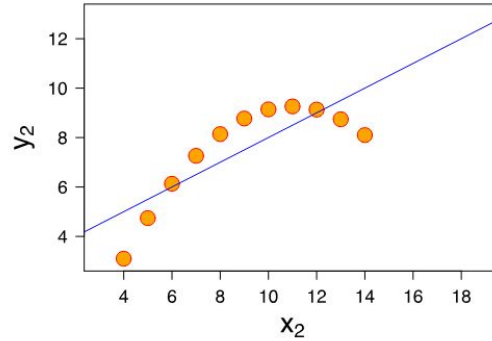
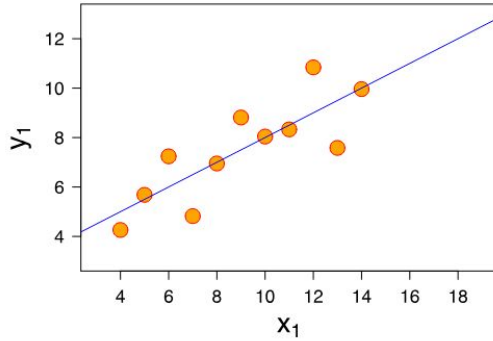




California housing



Visualize your data



4 data sets

Nearly identical statistics

Very different plots

source:

https://en.wikipedia.org/wiki/Anscombe%27s_quartet

California housing **tasks**

- Import & visualize the data (datasets/housing.csv)
- Split the data set to a training set and a test set (stratified, 70:30)
- Compute/visualize correlations ("median_house_value", "median_income", "total_rooms", "housing_median_age")
- Prepare the training set for ML algorithms:
 - Add new features
 - Impute features with missing values
 - Scale the data
- Learning:
 - Choose appropriate algorithms
 - Use internal cross-validation to tune the parameters
 - Evaluate on training set
- Evaluate on test set