Predictive Modelling with Python

Jure Žabkar

Contents

Software installation

Regression

Data preparation, Visualization, Modelling, Feature selection, Evaluation

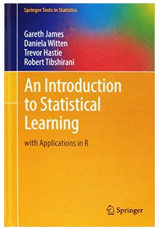
Classification

Data preparation, Visualization, Modelling, Feature selection, Evaluation

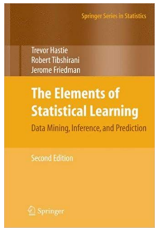
Installation

Install conda

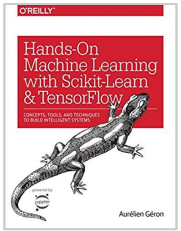
Readings



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

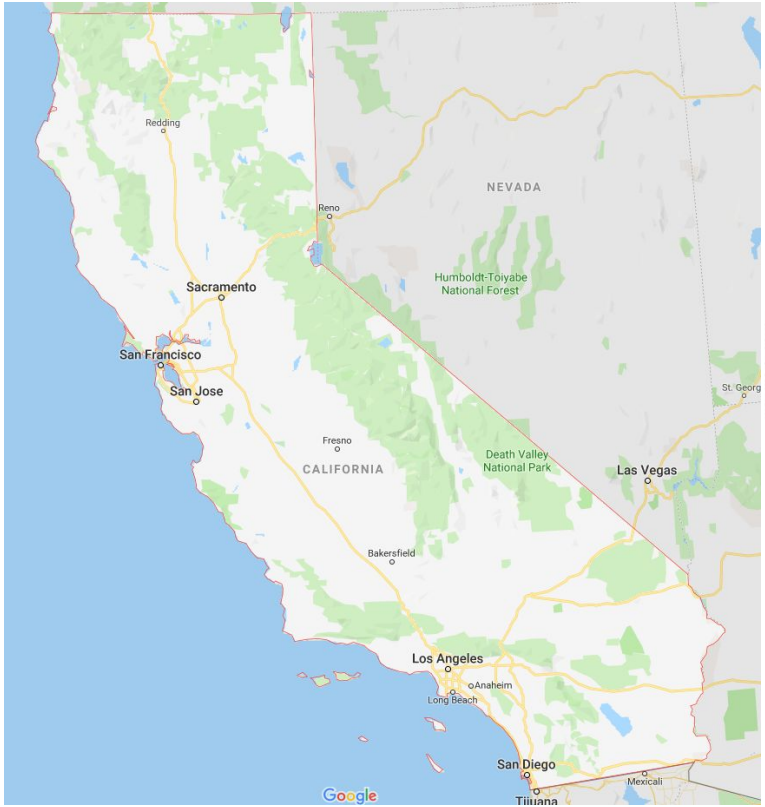


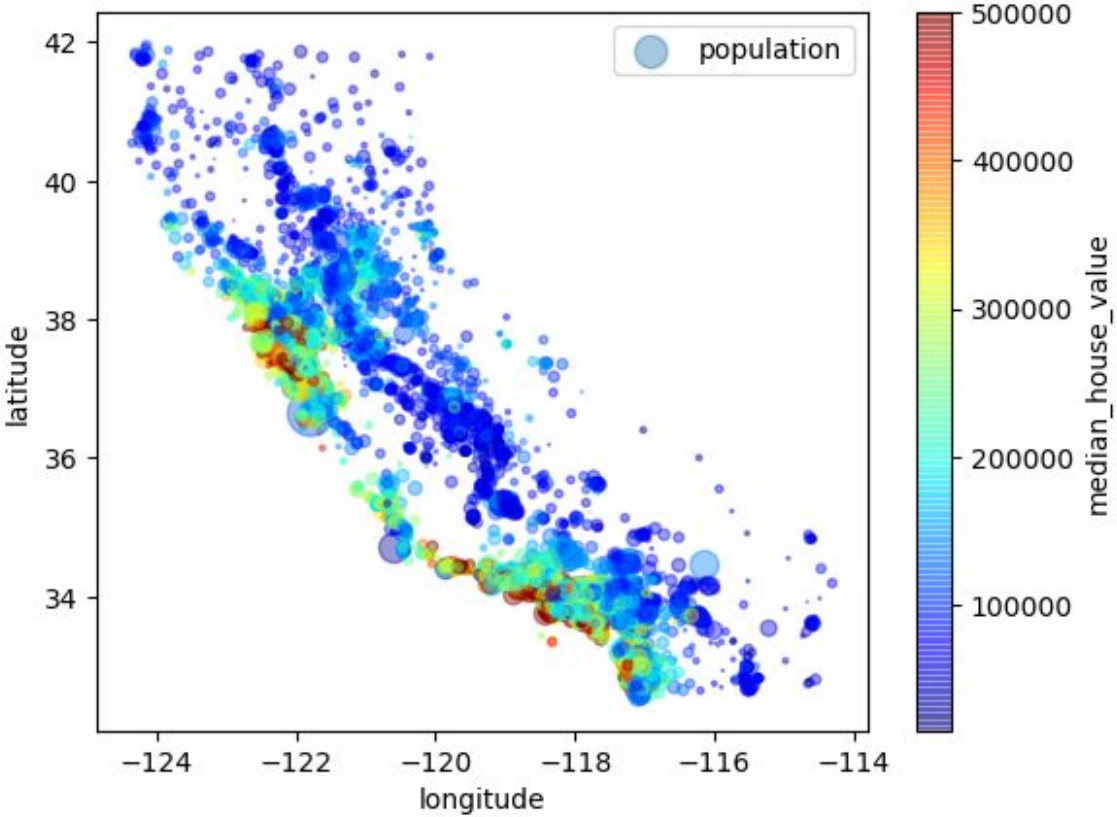
Friedman, J., Hastie, T., & Tibshirani, R. (2009). ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction***. Springer Series in Statistics.



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

California housing





Does **more money**

make **people happier?**

Classification

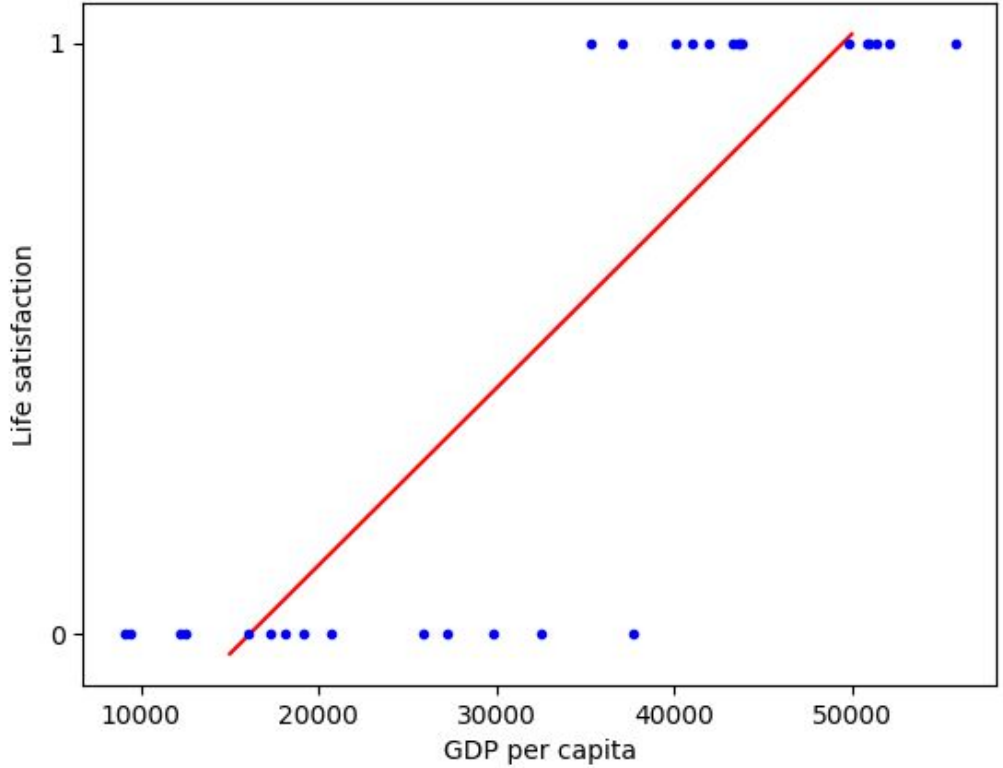
k Nearest Neighbour Classification

Logistic Regression

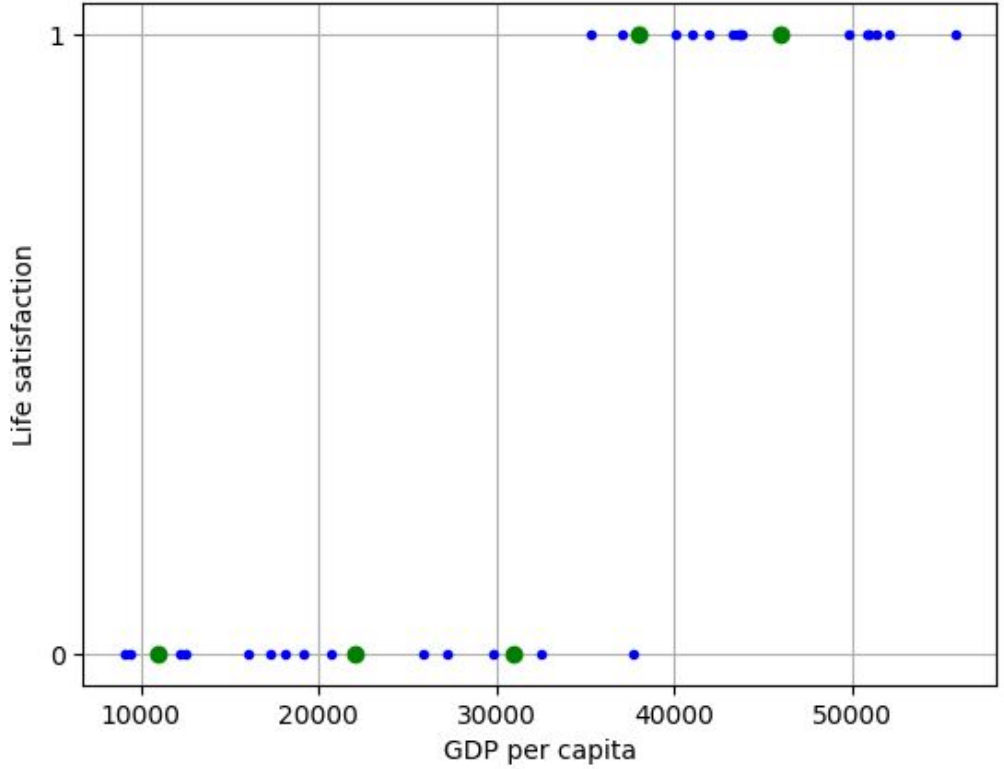
Classification Trees

Random Forest

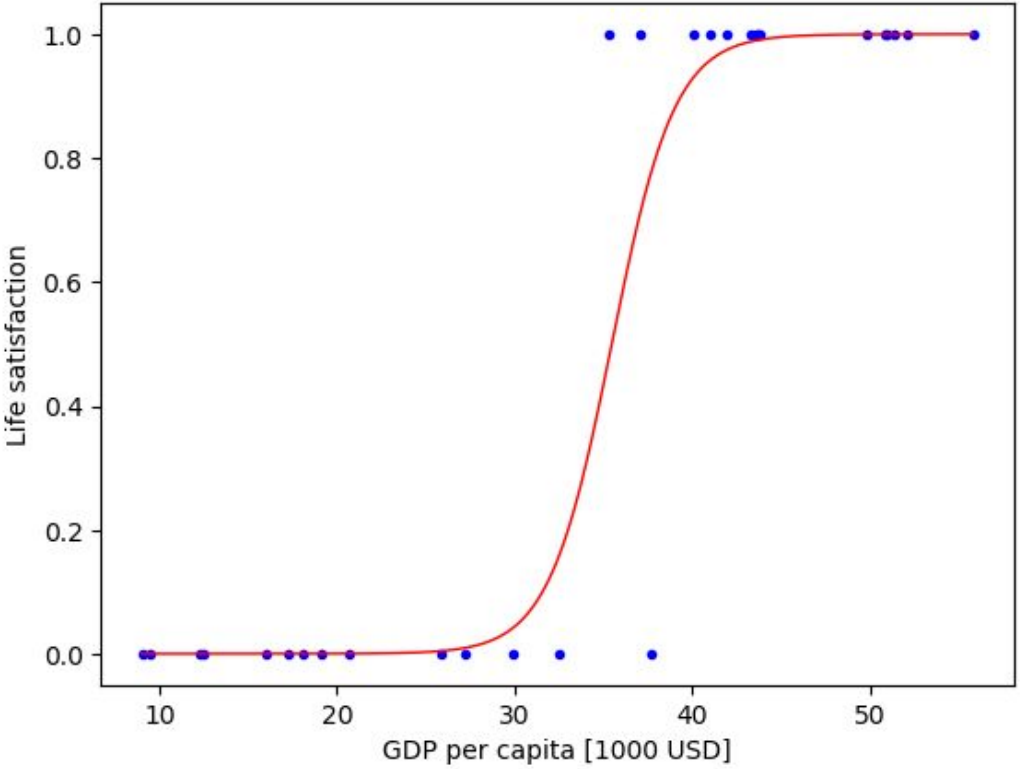
Life satisfaction linear regression



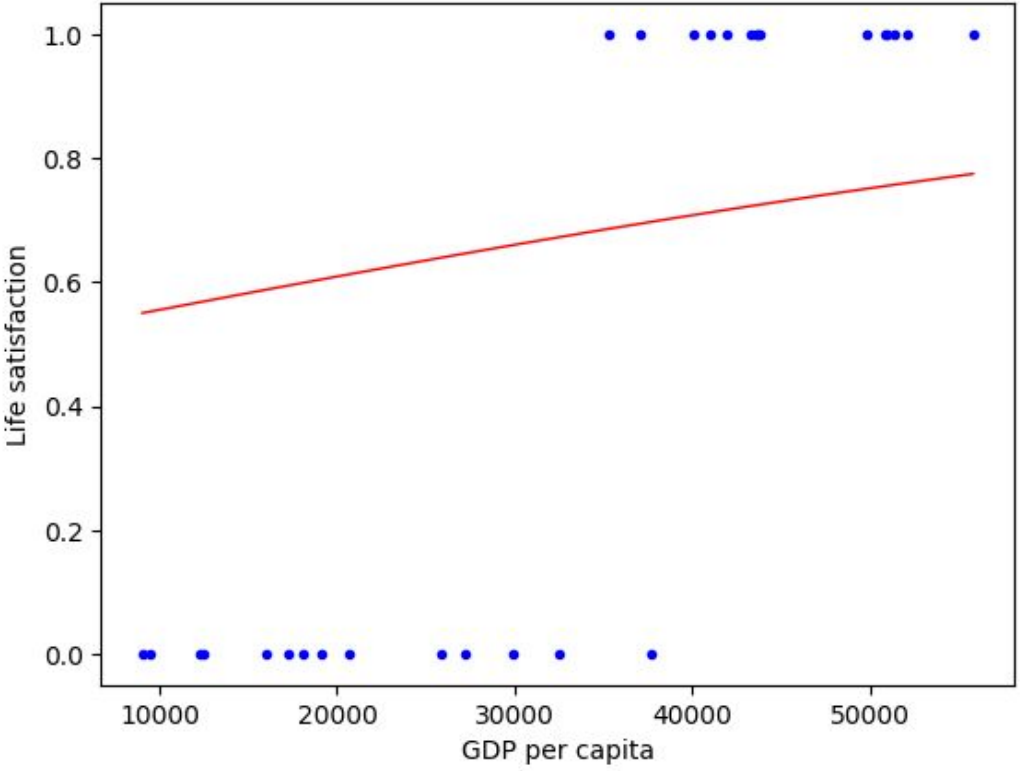
Life satisfaction kNN Classification



Life satisfaction **Logistic Regression**



Logistic Regression **Problems**



Evaluation

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negatives (TN)	False Positives (FP)
	Positive	False Negatives (FN)	True Positives (TP)

$$\text{precision} = \frac{TP}{TP + FP}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{TP}{TP + \frac{FN + FP}{2}}$$