

ITMAL Øvelser – Regression

Øvelse 1 :

I skal afprøve lineær regression på California Housing Dataset.

- Plot “median house value for California districts” (target værdi) i forhold til “MedInc median income in block” (første feature / input værdi).
- Prædikter hus-priser udfra kun 1 feature (fx. median income). Hvor stor er fejlen ? (root-mean-square fejl). Lav et plot af residual-værdierne som funktion af feature-værdi – ser der ud til at være sammenhæng mellem feature-værdi og residual ? Plot også et histogram over residualer – er de normalfordelte ?
- Prædikter hus-priser udfra alle 8 features. Find root mean squared error (RMSE).

Tips :

California Housing Dataset - <https://scikit-learn.org/dev/datasets/index.html#california-housing-dataset>

Lineær regression eksempel - https://scikit-learn.org/dev/auto_examples/linear_model/plot_ols.html#sphx-glr-auto-examples-linear-model-plot-ols-py

Øvelse 2 :

I skal afprøve logistisk regression på Height-Weight datasæt (højde og vægt på hhv. mænd og kvinder) – se “read_height_weight_dataset.py”.

- Først skal I afprøve på kun højde data. I skal plotte højde data for mænd (kategori 0) og kvinder (kategori 1) og dernæst fitte logistisk regression til data. I skal plotte den logistiske funktion.
- Dernæst skal I fitte til 2D data og plotte “højde-kurven” for den logistiske funktion (OBS: Når den er lig 0.5, har I decision boundary – dvs. begge klasser lige sandsynlige).
- Lav en rigtig træning med train/test sæt split. Find train og test error rates.

Tips :

Fit logistisk funktion - https://scikit-learn.org/dev/modules/generated/sklearn.linear_model.LogisticRegression.html#sklearn.linear_model.LogisticRegression

Contour plot - https://matplotlib.org/gallery/images_contours_and_fields/contour_demo.html#sphx-glr-gallery-images-contours-and-fields-contour-demo-py

General surface plot - https://matplotlib.org/mpl_toolkits/mplot3d/tutorial.html