# Viikko 39 -tehtävät

## Tehtävä 1

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sb

from sklearn.linear\_model import LinearRegression

x = [1, 2, 3, 4, 6, 7, 8]

y = [2\*x + 3 for x in x]

df = pd.DataFrame({'X': x, 'Y': y})

plt.figure(figsize=(8, 6))

plt.scatter(df['X'], df['Y'], color='blue')

plt.plot(df['X'], df['Y'], color='red')

plt.grid(True)

plt.show()

Kuva, joka sisältää kohteen viiva, Tontti, diagrammi, Samansuuntainen

Kuvaus luotu automaattisesti

## Tehtävä 2

model = LinearRegression()

model.fit(df[['X']], df['Y'])

coef = model.coef\_

inter = model.intercept\_

print(f'Suoran yhtälö: y = {coef[0]}x + {inter}')



plt.scatter(df['X'], df['Y'], color='blue')

plt.plot(df['X'], df['Y'], color='blue')

plt.scatter(x=5, y=model.predict([[5]]), color='red')

plt.grid(True)

plt.show()

Kuva, joka sisältää kohteen viiva, Tontti, diagrammi, numero

Kuvaus luotu automaattisesti

## Tehtävä 3

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from scipy.stats import pearsonr

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import r2\_score, mean\_absolute\_error, mean\_squared\_error

df = pd.read\_csv('./work/viikko5/datasets/salary.csv')

plt.scatter(df['YearsExperience'], df['Salary'])

plt.show()

Kuva, joka sisältää kohteen kuvakaappaus, diagrammi, teksti, viiva

Kuvaus luotu automaattisesti

corr = df.corr()

sns.heatmap(corr, annot=True)

Kuva, joka sisältää kohteen teksti, kuvakaappaus

Kuvaus luotu automaattisesti

pearsonr(df['YearsExperience'], df['Salary'])



X = df.iloc[:, [0]]

y = df.iloc[:, [1]]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.3, random\_state=0)

model = LinearRegression()

model.fit(X\_train, y\_train)

coef = model.coef\_

inter = model.intercept\_

print(f'Suoran yhtälö: Salary = {coef[0]} \* YearsOfExperience + {inter}')



y\_pred = model.predict(X\_test)

plt.scatter(X\_test, y\_test, color='red')

plt.plot(X\_test, y\_pred, color='blue')

plt.show()

Kuva, joka sisältää kohteen viiva, kuvakaappaus, teksti, Tontti

Kuvaus luotu automaattisesti

sns.regplot(x=X\_test, y=y\_test)

plt.show()

Kuva, joka sisältää kohteen kuvakaappaus, viiva, diagrammi, Tontti

Kuvaus luotu automaattisesti

r2 = r2\_score(y\_test, y\_pred)

mae = mean\_absolute\_error(y\_test, y\_pred)

rmse = mean\_squared\_error(y\_test, y\_pred, squared=False)

print(f'r2: {r2}')

print(f'mae: {mae}')

print(f'rmse: {rmse}')

Kuva, joka sisältää kohteen teksti, Fontti, kuvakaappaus

Kuvaus luotu automaattisesti

print(f'Palkka 7 vuoden kokemuksella: {model.predict([[7]])}')



## Tehtävä 4

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from scipy.stats import pearsonr

from sklearn.linear\_model import LinearRegression

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import r2\_score, mean\_absolute\_error, mean\_squared\_error

df = pd.read\_csv('./work/viikko5/datasets/housing.csv')

plt.scatter(x=df['median\_income'], y=df['median\_house\_value'], s=1)

plt.show()

Kuva, joka sisältää kohteen teksti, kuvakaappaus

Kuvaus luotu automaattisesti

X = df[['median\_income']]

y = df[['median\_house\_value']]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=0)

model = LinearRegression()

model.fit(X\_train, y\_train)

coef = model.coef\_

inter = model.intercept\_

print(f'Suoran yhtälö: MedianHouseValue = {coef[0]} \* MedianIncome + {inter}')



y\_pred = model.predict(X\_test)

plt.hist(y\_test - y\_pred, bins=50)

plt.xlabel('Ero todelliseen arvoon')

plt.ylabel('Lukumäärä')

plt.show()

Kuva, joka sisältää kohteen diagrammi, kuvakaappaus, teksti, Tontti

Kuvaus luotu automaattisesti

r2 = r2\_score(y\_test, y\_pred)

mae = mean\_absolute\_error(y\_test, y\_pred)

rmse = mean\_squared\_error(y\_test, y\_pred, squared=False)

print(f'r2: {r2}')

print(f'mae: {mae}')

print(f'rmse: {rmse}')

Kuva, joka sisältää kohteen teksti, Fontti, kuvakaappaus

Kuvaus luotu automaattisesti

print(f'Ennustettu talon arvo 30 000$ vuosituloilla: {model.predict([[30]])}')

