

QUIZ I COMP6577-MACHINE LEARNING 2022

1. A. Diketahui sample space dari hasil tossing 2 coins adalah $S = \{HH, HT, TH, TT\}$. Bila X adalah jumlah kemunculan heads maka diperoleh $X(HH) = 2$, $X(TH) = 1$, $X(HT) = 1$, and $X(TT) = 0$. Tentukan probability mass function $P(X)$ and hitung ekspektasi $E[X]$.

- B. Kovarian dari 2 variabel acak X dan Y didefinisikan sebagai

$$\text{cov}(x, y) := \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])]$$

Dan the corresponding korelasinya didefinisikan sebagai

$$r_{xy} := \mathbb{E}[xy] = \text{cov}(x, y) + \mathbb{E}[x] \mathbb{E}[y]$$

Hitung (by hand) kovarian dan korelasi berdasarkan random variable X dan Y serta distribusi gabungan P di bawah ini.

X	1	2	3	4
Y	0	1	1	2
P	0.2	0.3	0.2	0.3

2. Sdr diminta membantu rekan Sdr untuk mengestimasi harga sebuah rumah di suatu kota tertentu di US. Kepada Sdr diberikan dataset perumahan di kota tersebut yaitu "housing.csv" (<https://www.kaggle.com/code/prasadperera/the-boston-housing-dataset/data>) yang memuat informasi hubungan antara 13 variabel bebas (CRIM, ZN, dstnya) dengan variabel output MEDV (PRICE).

Untuk mendapatkan hasil prediksi yang akurat, pastikan data HOUSING.CSV yang akan Sdr gunakan adalah data yang sudah "cleaned" dari *missing values* dan *outlier*. Lebih lanjut, pastikan juga Sdr hanya memilih variabel bebas yang memiliki "strong correlation" terhadap variabel independent. Gunakan teknik filtering untuk memilih variabel bebas yang tepat. Implementasikan model Sdr dalam python notebook untuk proses-proses antara lain feature engineering, model training dan validation, serta model evaluation. PLS USE GRADIENT_DESCENT FUNCTION TAKEN FROM YOUR 1ST ASSIGNMENT.

3. Sdr diminta membuat sebuah model Regresi yang dapat membantu memutuskan apakah calon mahasiswa yang mendaftar dapat diterima atau ditolak berdasarkan dataset "Mark.txt" yang memuat nilai test 1 dan test 2. Split dataset utk pelatihan dan uji validasi. Implementasikan model Sdr dalam python notebook dan evaluasi kinerja model regresi tersebut dengan confusion matrix CM. Pastikan Sdr telah melakukan standarisasi data sebelum melakukan proses pelatihan dan validasi. PLS USE GRADIENT_DESCENT FUNCTION TAKEN FROM YOUR 2ND ASSIGNMENT (LOG REGRESSION).
4. Bandingkan hasil Sdr pada problem no. 2 di atas bila Sdr menggunakan model Naïve Bayes classifier. Silahkan gunakan pustaka yang relevant yang disediakan oleh *scikitlearn python*.