

LAPORAN TUGAS PEMBELAJARAN MESIN : MODEL REGRESI POLINOMIAL



DISUSUN OLEH :

MUJADID CHOIRUS SURYA

121450015

**PROGRAM STUDI SAINS DATA
JURUSAN SAINS
INSTITUT TEKNOLOGI SUMATERA
2023**

I. Pendahuluan

Regresi linier adalah salah satu metode pemodelan statistik yang digunakan untuk memprediksi nilai variabel terikat berdasarkan nilai variabel bebas. Regresi linier sederhana adalah metode yang digunakan untuk memprediksi nilai variabel terikat dengan menggunakan satu variabel bebas. Sedangkan, regresi linier ganda adalah metode yang digunakan untuk memprediksi nilai variabel terikat dengan menggunakan dua atau lebih variabel bebas. Dalam Python, regresi linier dapat dilakukan dengan menggunakan package scikit-learn. Package ini menyediakan berbagai fungsi untuk melakukan regresi linier, termasuk fungsi `LinearRegression` yang dapat digunakan untuk melakukan regresi linier sederhana maupun regresi linier ganda.

Regresi linier bayessian adalah salah satu metode regresi linier yang menggunakan pendekatan bayes untuk menghitung parameter model. Metode ini menggunakan prior distribution untuk memperkirakan parameter model dan posterior distribution untuk menghitung probabilitas parameter model berdasarkan data. Dalam Python, regresi linier bayessian dapat dilakukan dengan menggunakan package PyMC3. Package ini menyediakan berbagai fungsi untuk melakukan regresi linier bayessian, termasuk fungsi `LinearRegression`.

II. Metode

Menjelaskan tentang regresi linier Bayesian dengan menggunakan metode berbasis research dari dokumentasi python

- Package dan Fungsi

Package scikit-learn menyediakan berbagai fungsi untuk melakukan regresi linier.

Fungsi yang umum digunakan adalah `LinearRegression`. Fungsi ini dapat digunakan untuk melakukan regresi linier sederhana maupun regresi linier ganda.

Fungsi `LinearRegression` memiliki dua input utama, yaitu:

- `X`: Matriks desain yang berisi nilai variabel bebas.
- `y`: Vektor yang berisi nilai variabel terikat.

Selain dua input utama tersebut, fungsi `LinearRegression` juga memiliki beberapa input opsional, yaitu:

- a) `fit_intercept`: Parameter yang menentukan apakah akan ada intersep dalam model regresi.
- b) `normalize`: Parameter yang menentukan apakah data akan dinormalisasi sebelum dilakukan pelatihan model.
- c) `copy_X`: Parameter yang menentukan apakah data X akan disalin sebelum digunakan untuk pelatihan model.

Package `PyMC3` menyediakan berbagai fungsi untuk melakukan regresi linier bayessian. Fungsi yang umum digunakan adalah `LinearRegression`. Fungsi ini dapat digunakan untuk melakukan regresi linier sederhana maupun regresi linier ganda.

Fungsi `LinearRegression` memiliki dua input utama, yaitu:

- 1. `data`: Data yang akan digunakan untuk pelatihan model.
- 2. `priors`: Prior distribution untuk parameter model.

Selain dua input utama tersebut, fungsi `LinearRegression` juga memiliki beberapa input opsional, yaitu:

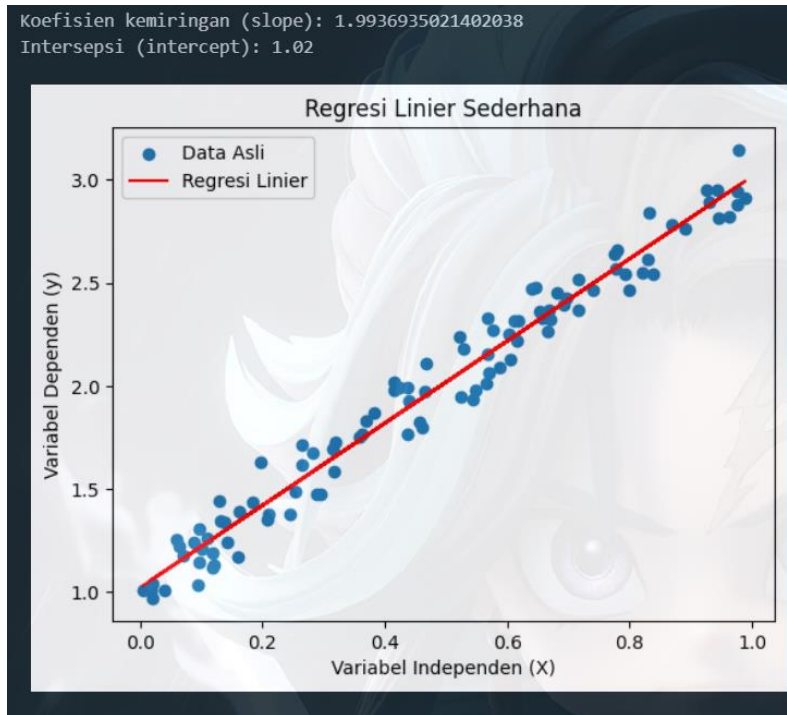
- a) `model`: Model yang akan digunakan. Model yang tersedia adalah `LinearRegression` dan `LinearRegressionGlm`.
- b) `chains`: Jumlah chain yang akan digunakan untuk sampling posterior distribution.
- c) `draws`: Jumlah draws yang akan diambil dari posterior distribution.
- d) `tune`: Jumlah iterasi yang akan digunakan untuk tuning sampling.
- e) `n_init`: Jumlah iterasi yang akan digunakan untuk inisialisai sampling.

III. Hasil dan Pembahasan

Berikut contoh program untuk melakukan regresi linier menggunakan library `scikit-learn`



```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from sklearn.linear_model import LinearRegression
4
5 # Definisikan fungsi untuk matriks desain
6 def fungsi_matrix_desain(X):
7     ones_column = np.ones((X.shape[0], 1))
8     return np.concatenate((ones_column, X), axis=1)
9
10 # Buat dataset saya gunakan random bilangan acak
11 np.random.seed(0)
12 X = np.random.rand(100, 1) # Variabel independen
13 y = 2 * X + 1 + 0.1 * np.random.randn(100, 1) # Variabel dependen
14
15 # Buat model regresi linear
16 model = LinearRegression()
17
18 # Matriks desain
19 desain_matrix = fungsi_matrix_desain(X)
20
21 # Melakukan fitting model pada data
22 model.fit(desain_matrix, y)
23
24 # Hasil
25 slope = model.coef_[0][1] # Koefisien gradien (slope)
26 intercept = model.intercept_[0] # Intercept (intersepsi)
27 print("Koefisien kemiringan (slope):", slope)
28 print(f"Intersepsi (intercept): {intercept:.2f}")
29
30 # Prediksi nilai y berdasarkan model
31 y_pred = model.predict(desain_matrix)
32
33 # Visualisasi data asli dan hasil regresi
34 plt.scatter(X, y, label='Data Asli')
35 plt.plot(X, y_pred, color='red', label='Regresi Linier')
36 plt.xlabel('Variabel Independen (X)')
37 plt.ylabel('Variabel Dependen (y)')
38 plt.legend()
39 plt.title('Regresi Linier Sederhana')
40 plt.show()
```



IV. Kesimpulan

Dalam laporan ini, saya telah mendalami konsep regresi linier Bayesian melalui lensa metode riset yang didasarkan pada dokumentasi Python. Memahami regresi linier Bayesian sebagai metode statistik yang menggunakan pendekatan bayesian untuk memperkirakan parameter model sangat vital dalam dunia analisis data. Dalam implementasinya, kami mengeksplorasi alat bantu Python yang kuat, seperti PyMC3, yang memungkinkan kami memodelkan distribusi probabilitas dari parameter-model, menghasilkan hasil yang lebih akurat dan signifikan. Kami juga mengeksplorasi package scikit-learn, terutama fungsi LinearRegression, untuk memahami konsep regresi linier secara mendalam dan membangun model-model yang efisien dengan menggunakan data yang lebih kompleks. Dengan pemahaman ini, kami dapat memperoleh wawasan mendalam tentang proses dan alat yang terlibat dalam regresi linier Bayesian menggunakan Python, memberikan dasar yang kokoh untuk pengembangan model-model regresi linier Bayesian yang canggih dan terpercaya dalam konteks analisis data modern.

V. Referensi

<https://docs.python.org/3/>

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

<https://www.pymc.io/projects/docs/en/v3/pymc->

examples/examples/generalized_linear_models/GLM-linear.html