



الانحدار الخطي (تنفيذ بايثون)

[يقرأ](#)
[مناقشة\(20\)](#)
[الدورات](#)

تتناول هذه المقالة أساسيات الانحدار الخطي وتنفيذها في لغة البرمجة بايثون. الانحدار الخطي هو طريقة إحصائية لنموذج العلاقات بين متغير تابع مع مجموعة معينة من المتغيرات المستقلة.

ملحوظة: في هذه المقالة، نشير إلى المتغيرات التابعة كاستجابات والمتغيرات المستقلة كميزات للبساطة. من أجل توفير فهم أساسي للانحدار الخطي، نبدأ بالإصدار الأساسي من الانحدار الخطي، أي **الانحدار الخطي البسيط**.

ما هو الانحدار الخطي؟

الانحدار الخطي هو أسلوب إحصائي يستخدم للتنبؤ بمتغير تابع مستمر (المتغير المستهدف) بناءً على واحد أو أكثر من المتغيرات المستقلة (متغيرات التوقع). تفترض هذه التقنية وجود علاقة خطية بين المتغيرات التابعة والمستقلة، مما يعني أن المتغير التابع يتغير بشكل مناسب مع التغيرات في المتغيرات المستقلة. بمعنى آخر، يتم استخدام الانحدار الخطي لتحديد مدى قدرة متغير واحد أو أكثر على التنبؤ بقيمة المتغير التابع.

الافتراضات التي تقوم بها في نموذج الانحدار الخطي

فيما يلي الافتراضات الأساسية التي يضعها نموذج الانحدار الخطي فيما يتعلق بمجموعة البيانات التي يتم تطبيقه عليها:

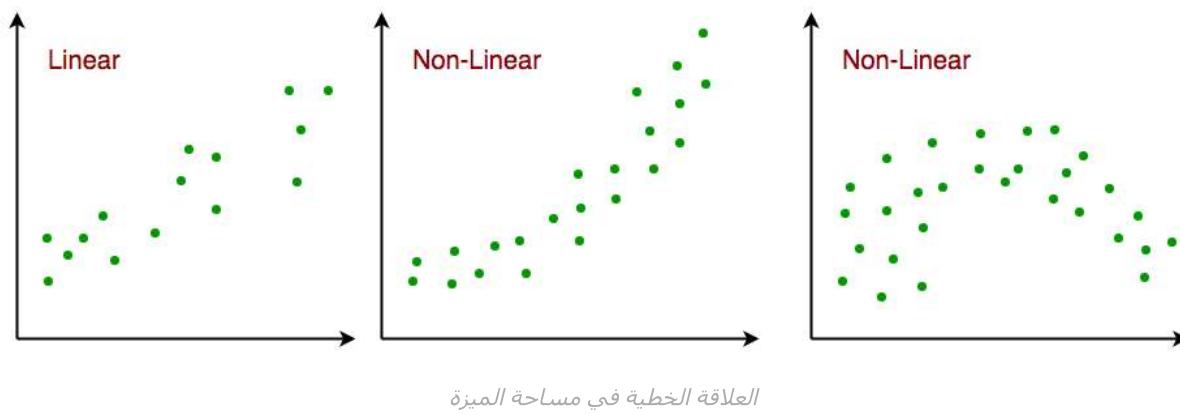


نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت [سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

فهمتها!

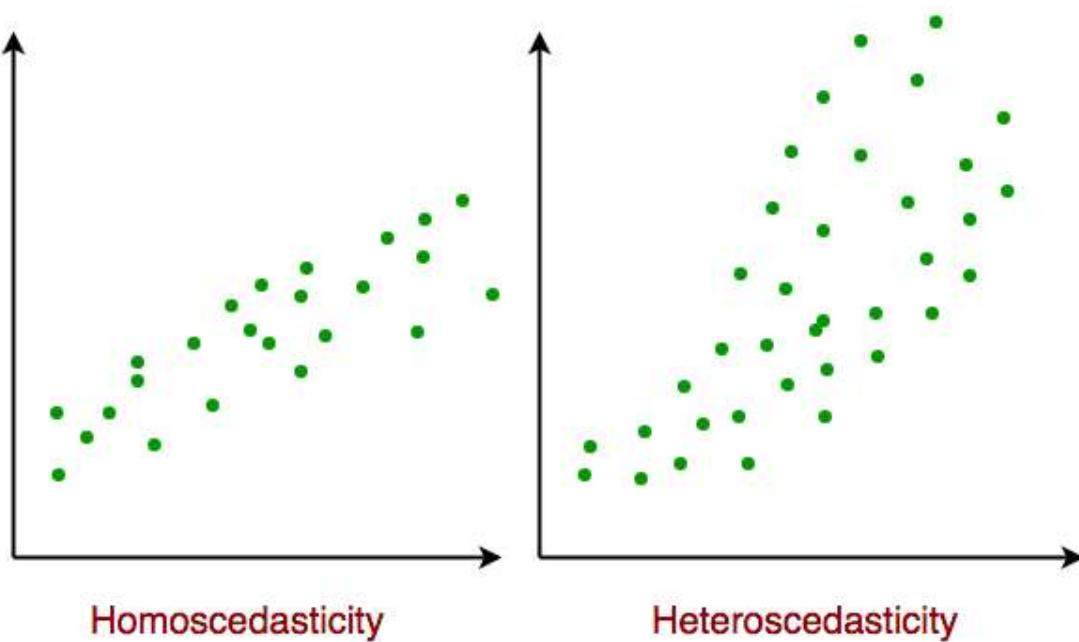
Spotlight

- العلاقة الخطية :** يجب أن تكون العلاقة بين الاستجابة ومتغيرات الميزة خطية. يمكن اختبار افتراض الخطية باستخدام المخططات المبعثرة. كما هو موضح أدناه، يمثل الشكل الأول المتغيرات المرتبطة خطياً في حين أن المتغيرات في السكلين الثاني والثالث على الأرجح غير خطية. لذا، فإن الشكل الأول سيعطي تنبؤات أفضل باستخدام الانحدار الخطى.



- وجود علاقة خطية متداخلة قليلة أو معدومة :** من المفترض أن هناك علاقة خطية متداخلة قليلة أو معدومة في البيانات. تحدث العلاقة الخطية المتعددة عندما لا تكون الميزات (أو المتغيرات المستقلة) مستقلة عن بعضها البعض.
- ارتباط ذاتي ضئيل أو معدوم :** هناك افتراض آخر وهو وجود ارتباط ذاتي ضئيل أو معدوم في البيانات. يحدث الارتباط التلقائي عندما لا تكون الأخطاء المتبقية مستقلة عن بعضها البعض. يمكنك الرجوع هنا لمزيد من المعرفة حول هذا الموضوع.
- لا توجد قيم متطرفة:** نحن نفترض أنه لا توجد قيم متطرفة في البيانات. القيم المتطرفة هي نقاط بيانات بعيدة عن بقية البيانات. يمكن أن تؤثر القيم المتطرفة على نتائج التحليل.
- المثلية :** تصف المثلية الحالة التي يكون فيها مصطلح الخطأ (أي "الضوضاء") أو الاضطراب العشوائي في العلاقة بين المتغيرات المستقلة والمتغير التابع) هو نفسه عبر جميع قيم المتغيرات المستقلة. كما هو موضح

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت [سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)



المتباينة في الانحدار الخطى

ومع وصولنا إلى نهاية هذه المقالة، سنناقش بعض تطبيقات الانحدار الخطى أدناه.

أنواع الانحدار الخطى

هناك نوعان رئيسيان من الانحدار الخطى:

- **الانحدار الخطى البسيط:** يتضمن التنبؤ بمتغير تابع بناءً على متغير مستقل واحد.
- **الانحدار الخطى المتعدد:** يتضمن التنبؤ بمتغير تابع بناءً على متغيرات مستقلة متعددة.

الانحدار الخطى البسيط

الانحدار الخطى البسيط هو أسلوب للتنبؤ بالاستجابة باستخدام ميزة واحدة. إنه أحد نماذج التعلم الآلي الأساسية التي يُعرف عليها عشاق التعلم الآلي. في الانحدار الخطى، نفترض أن المتغيرين، أي المتغيرات التابعية والمستقلة، بأكبر قدر ممكن من الدقة (y) مرتبان خطيا. ومن ثم، فإننا نحاول العثور على دالة خطية تتنبأ بقيمة الاستجابة y لكل ميزة x دعونا نفك في مجموعة بيانات حيث لدينا قيمة الاستجابة (y) كدالة للميزة أو المتغير المستقل (x):

x	0	1	2	3	4	5	6	7	8	9
y	1	3	2	5	7	8	8	9	10	12

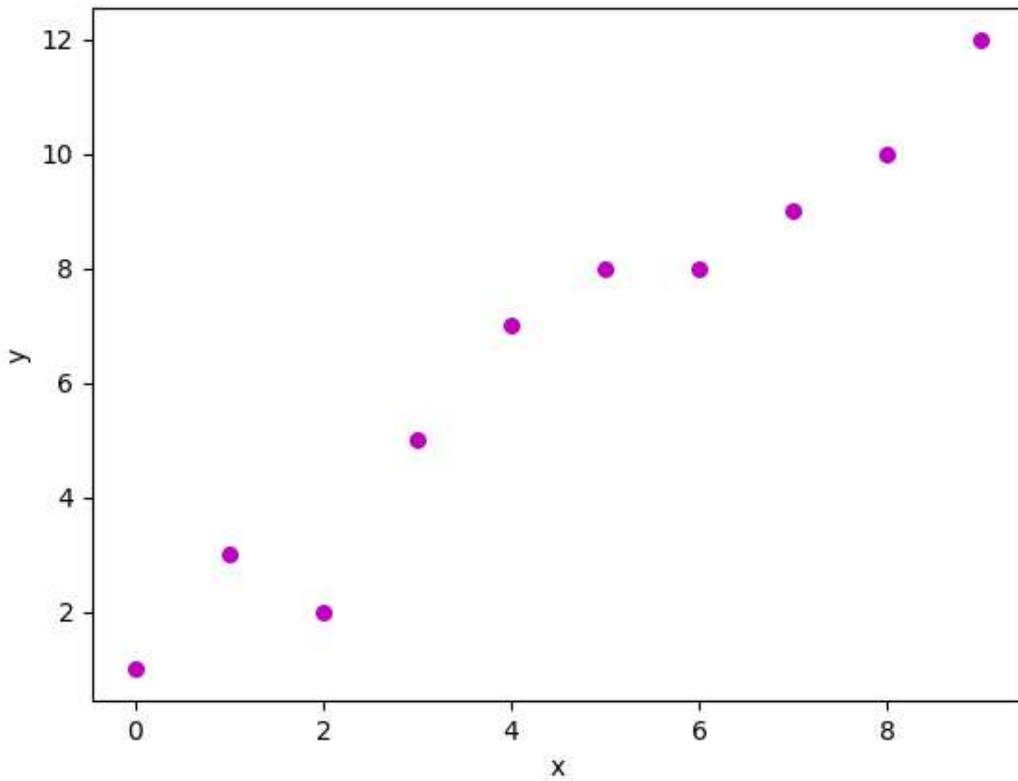
نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

وللعمومية نحدد:

x كمتجہ للمیزات ، أي $x = [x_1, x_2, \dots, x_n]$,

y كمتجہ للاستجابة ، أي $y = [y_1, y_2, \dots, y_n]$

- يبدو المخطط المبعثر لمجموعة البيانات أعلاه كما يلي . في المثال أعلاه $n=10$. للاحظات



مؤامرة مبعثر للبيانات التي تم إنشاؤها عشوائيا

الآن، المهمة هي العثور على الخط الذي يناسب بشكل أفضل المخطط المبعثر أعلاه حتى نتمكن من التنبؤ بـ **هذا الخط بخط الانحدار** . (غير موجودة في مجموعة بيانات x أي قيمة) . بالاستجابة لأي قيم ميزات جديدة يتم تمثيل معادلة خط الانحدار على النحو التالي :

$$h(x_i) = \beta_0 + \beta_1 x_i$$

، هنا

- **قيمة الاستجابة المتوقعة للملاحظة الأولى** $h(x_i)$ تمثل .

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

وبمجرد تقدير هذه المعاملات، b_1 و b_0 لإنشاء نموذجنا، يجب علينا "التعلم" أو تقدير قيم معاملات الانحدار يمكننا استخدام النموذج للتنبؤ بالاستجابات في هذه المقالة، سوف نستخدم مبدأ **المربعات الصغرى**.

:الآن فكر في

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = h(x_i) + \varepsilon_i \Rightarrow \varepsilon_i = y_i - h(x_i)$$

هو خطأ متبقى في الملاحظة. لذا، هدفنا هو تقليل إجمالي الأخطاء المتبقية. نحدد الخطأ التربيعي أو دالة e_i ، هنا على النحو التالي L ، التكالفة:

$$J(\beta_0, \beta_1) = \frac{1}{2n} \sum_{i=1}^n \varepsilon_i^2$$

لها الحد الأدنى! وبدون الدخول في التفاصيل (b_0, b_1) التي تكون b_0 و b_1 مهمتنا هي إيجاد قيمة الرياضية نعرض النتيجة هنا:

$$\beta_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

γ هو مجموع الانحرافات المتقاطعة SS_{xy} حيث:

$$SS_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n y_i x_i - n \bar{x} \bar{y}$$

\times هو مجموع انحرافات SS_{xx} و \times هو مجموع المربعة SS_{yy} :

$$SS_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n(\bar{x})^2$$

تنفيذ بايثون للانحدار الخطى البسيط

يمكننا استخدام لغة بايثون لمعرفة معامل نماذج الانحدار الخطى. لتخطيط بيانات الإدخال والخط الأفضل سنستخدم إنها واحدة من مكتبات بايثون الأكثر استخداماً لرسم الرسوم البيانية. هنا مثال على الانحدار . matplotlib مكتبة simple باستخدام بايثون الخطى.

استيراد المكتبات

Python3

```
import numpy as np
import matplotlib.pyplot as plt
```

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
سياسة ملفات تعريف الارتباط وسياسة الخصوصية الخاصة بنا

وتقدير معاملات (المتغير التابع) y و (المتغير المستقل) x بأخذ البيانات المدخلة(`estimate_coef()`) تقوم هذه الدالة خط الانحدار الخطى باستخدام طريقة المربعات الصغرى.

- حساب عدد الملاحظات: يحدد عدد نقاط البيانات $n = np.size(x)$
- حساب الوسائل: وحساب القيم المتوسطة $m_x = np.mean(x)$ و $m_y = np.mean(y)$
- وحساب مجموع الانحرافات المربعة بين و مجموع: x حساب الانحراف العرضي والانحراف حول الانحرافات المربعة لمتوسطها على التوالى $ss_xy = np.sum(y*x) - n*m_y*m_x$
 $ss_xx = np.sum(x*x) - n*m_x*m_xx$
- حساب معاملات الانحدار: وتحديد ميل () وتقاطع () خط الانحدار باستخدام طريقة المربعات الصغرى
 $b_1 = ss_xy / ss_xx$
 $b_0 = m_y - b_1 * m_x$
- إرجاع المعاملات: تقوم الدالة بإرجاع المعاملات المقدرة على شكل صف (b_0, b_1) .

Python3

```
def estimate_coef(x, y):
    # number of observations/points
    n = np.size(x)

    # mean of x and y vector
    m_x = np.mean(x)
    m_y = np.mean(y)

    # calculating cross-deviation and deviation about x
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x

    # calculating regression coefficients
    b_1 = SS_xy / SS_xx
    b_0 = m_y - b_1*m_x

    return (b_0, b_1)
```

رسم وظيفة خط الانحدار

والمعاملات ، (متغير تابع) y ، (متغير مستقل) x بيانات الإدخال(`plot_regression_line()`) تأخذ هذه الوظيفة لرسم خط الانحدار ونقطة البيانات b المقدرة.

1. رسم مخطط مبعثر: يرسم نقاط البيانات الأصلية كمخطط مبعثر بعلامات حمراء `plt.scatter(x, y,`

`marker="o", color="red")`

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك توافق بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و [سياسة الخصوصية الخاصة بنا](#)

2. حساب متوجه الاستجابة المتوقعة: حساب القيم المتوقعة بناءً على المعاملات المقدرة $y_{pred} = b[0] + b[1]*xyb$
3. رسم خط الانحدار: رسم خط الانحدار باستخدام القيم المتوقعة والمتغير المستقل `plt.plot(x, y_pred, color = "g")`
4. إضافة تسميات: وقم بتسمية المحور السيني والمحور الصادي ك على التوالي `plt.xlabel('x') plt.ylabel('y')`
-

Python3

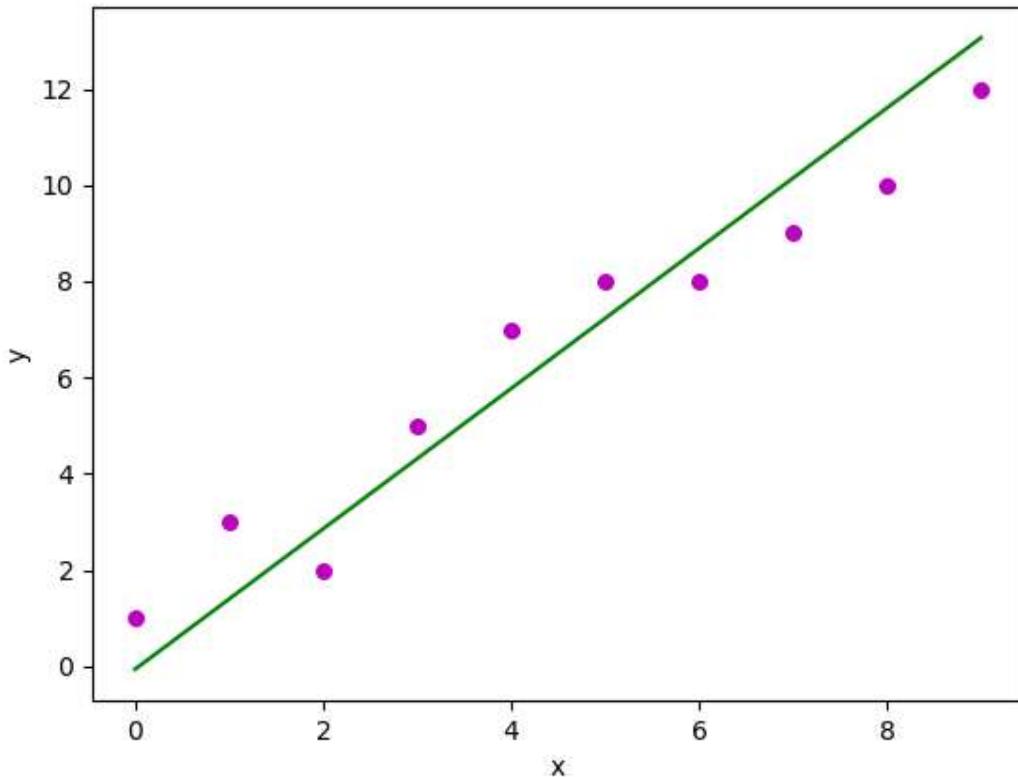
```
def plot_regression_line(x, y, b):
    # plotting the actual points as scatter plot
    plt.scatter(x, y, color = "m",
                marker = "o", s = 30)

    # predicted response vector
    y_pred = b[0] + b[1]*x

    # plotting the regression line
    plt.plot(x, y_pred, color = "g")

    # putting labels
    plt.xlabel('x')
    plt.ylabel('y')
```

انتاج:



مخطط التشتت لل نقاط مع خط الانحدار

الوظيفة الأساسية

يقوم بالخطوات التالية() main يطبق الكود المقدم تحليل الانحدار الخطى البسيط من خلال تحديد دالة

- تعريف البيانات: يحدد المتغير المستقل 1. كمصفوفات (y) والمتغير التابع (x) NumPy.

- تقدير المعامل: يستدعي الدالة 2. لتحديد معاملات خط الانحدار الخطى باستخدام البيانات estimate_coef() المقدمة.

- لخط الانحدار (b_1) والمنحدر (b_0) 3. معاملات الطباعة: طباعة التقاطع المقدر.

Python3

```
def main():
    # observations / data
    x = np.array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
    y = np.array([1, 3, 2, 5, 7, 8, 8, 9, 10, 12])

    # estimating coefficients
    b = estimate_coef(x, y)
    print("Estimated coefficients:\nb_0 = {} \
          \nb_1 = {}".format(b))
```

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

انتاج:

المعاملات المقدرة :

$$\underline{b}_0 = 0.0586206896552$$

$$\underline{b}_1 = 1.45747126437$$

الانحدار الخطى المتعدد

يحاول الانحدار الخطى المتعدد وضع نموذج للعلاقة بين ميزتين أو أكثر والاستجابة عن طريق تركيب معادلة خطية على البيانات المرصودة.

أو p ومن الواضح أنه ليس سوى امتداد للانحدار الخطى البسيط. فكر في مجموعة بيانات تحتوي على ميزات واستجابة واحدة (أو متغير تابع) (متغيرات مستقلة

من الصنوف/الملاحظات n تحتوي مجموعة البيانات أيضاً على عدد

نحدد:

أ. للمرأة j إلى قيم الميزة \underline{x}_j حيث تشير $p \times n$ مصفوفة بالحجم = (مصفوفة السمات) \underline{X} .

لذا،

$$\begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \vdots & x_{np} \end{pmatrix}$$

و

أ. إلى قيمة الاستجابة للملاحظة $\{i\}$ حيث تشير n متوجه الحجم = (متوجه الاستجابة) \underline{y} .

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

على النحو التالي p يتم تمثيل خط الانحدار لميزات

$$h(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

هي معاملات الانحدار . وأيضاً b_0, b_1, \dots, b_p هي قيمة الاستجابة المتوقعة للملاحظة (i) حيث $h(x_i)$

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و [سياسة الخصوصية الخاصة بنا](#)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$$

or

$$y_i = h(x_i) + \varepsilon_i \Rightarrow \varepsilon_i = y_i - h(x_i)$$

خطأً متبقياً في الملاحظة الأولى. يمكننا تعليم نموذجنا الخطى أكثر قليلاً من خلال تمثيل مصفوفة ε حيث يمثل على النحو التالي X الميزة:

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

والآن، يمكن التعبير عن النموذج الخطى بدالة المصفوفات على النحو التالي:

$$y = X\beta + \varepsilon$$

أين،

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}$$

و

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

باستخدام طريقة المربعات الصغرى . كما هو موضح سابقاً، تمثل طريقة المربعات 'b' أي ، b الآن، نحدد تقديرًا له الذي يتم من خلاله تقليل إجمالي الخطأ المتبقى 'b' الصغرى إلى تحديد ونعرض النتيجة مباشرة هنا:

$$\hat{\beta} = (X'X)^{-1}X'y$$

يمكن ، 'b' ، حيث يمثل 'b' تبديل المصفوفة بينما يمثل - 1 مكوس المصفوفة . بمعرفة تقديرات المربيعات الصغرى الآن تقدير نموذج الانحدار الخطى المتعدد على النحو التالي:

$$\hat{y} = X\hat{\beta}$$

. هو متجه الاستجابة المقدرة 'y' حيث

تنفيذ بايثون للانحدار الخطى المتعدد

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت سياسة ملفات تعريف الارتباط وسياسة الخصوصية الخاصة بنا

Python3

```
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets, linear_model, metrics
```

تحميل مجموعة بيانات الإسكان في بوسطن

المقدم ويقرأها في URL من عنوان Boston Housing يقوم الكود بتنزيل مجموعة بيانات DataFrame (raw_df)

Python3

```
data_url = "http://lib.stat.cmu.edu/datasets/boston"
raw_df = pd.read_csv(data_url, sep="\s+",
                     skiprows=22, header=None)
```

معالجة البيانات

يتم تحديد متغيرات DataFrame من (y) والمتغير الهدف (X) يؤدي هذا إلى استخراج متغيرات الإدخال الإدخال من كل صف آخر لمطابقة المتغير المستهدف، والذي يتوفّر في كل صف آخر.

Python3

```
X = np.hstack([raw_df.values[:, :-2],
               raw_df.values[:, -2:]])
y = raw_df.values[:, -2]
```

تقسيم البيانات إلى مجموعات التدريب والاختبار

الوظيفة من train_test_split() هنا يقوم بتقسيم البيانات إلى مجموعات تدريب واختبار باستخدام scikit-learn. أنه يجب استخدام 40% من البيانات لاختبار test_size تحدد المعلمة.

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت سياسة ملفات تعريف الارتباط وسياسة الخصوصية الخاصة بنا

```
x_train, x_test,\n    y_train, y_test = train_test_split(X, y,\n                                         test_size=0.4,\n                                         random_state=1)
```

إنشاء وتدريب نموذج الانحدار الخطى

يؤدي هذا إلى تهيئة كائن `LinearRegression` (`reg`) وتدريب النموذج باستخدام بيانات التدريب (`x_train` ، `y_train`)

Python3

```
reg = linear_model.LinearRegression()\nreg.fit(x_train, y_train)
```

تقييم أداء النموذج

يقوم بتقييم أداء النموذج من خلال طباعة معاملات الانحدار وحساب درجة التباين، والتي تقيس نسبة التباين الموضح. تشير النتيجة 1 إلى التنبؤ المثالي.

Python3

```
# regression coefficients\nprint('Coefficients: ', reg.coef_)\n\n# variance score: 1 means perfect prediction\nprint('Variance score: {}'.format(reg.score(x_test, y_test)))
```

انتاج:

:المعاملات

```
[ -8.80740828e-02 6.72507352e-02 5.10280463e-02 2.18879172e+00\n-1.72283734e+01 3.62985243e+00 2.13933641e-03 -1.36531300e+00\n2.88788067e-01 -1.22618657e-02 -8.36014969e-01 9.53058061e-03
```

نحو نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

-5.05036163e-01]

درجة التباين: 0.720898784611

رسم الأخطاء المتبقية

رسم وتحليل الأخطاء المتبقية والتي تمثل الفرق بين القيم المتوقعة والقيم الفعلية.

Python3

```
# plot for residual error

# setting plot style
plt.style.use('fivethirtyeight')

# plotting residual errors in training data
plt.scatter(reg.predict(X_train),
            reg.predict(X_train) - y_train,
            color="green", s=10,
            label='Train data')

# plotting residual errors in test data
plt.scatter(reg.predict(X_test),
            reg.predict(X_test) - y_test,
            color="blue", s=10,
            label='Test data')

# plotting line for zero residual error
plt.hlines(y=0, xmin=0, xmax=50, linewidth=2)

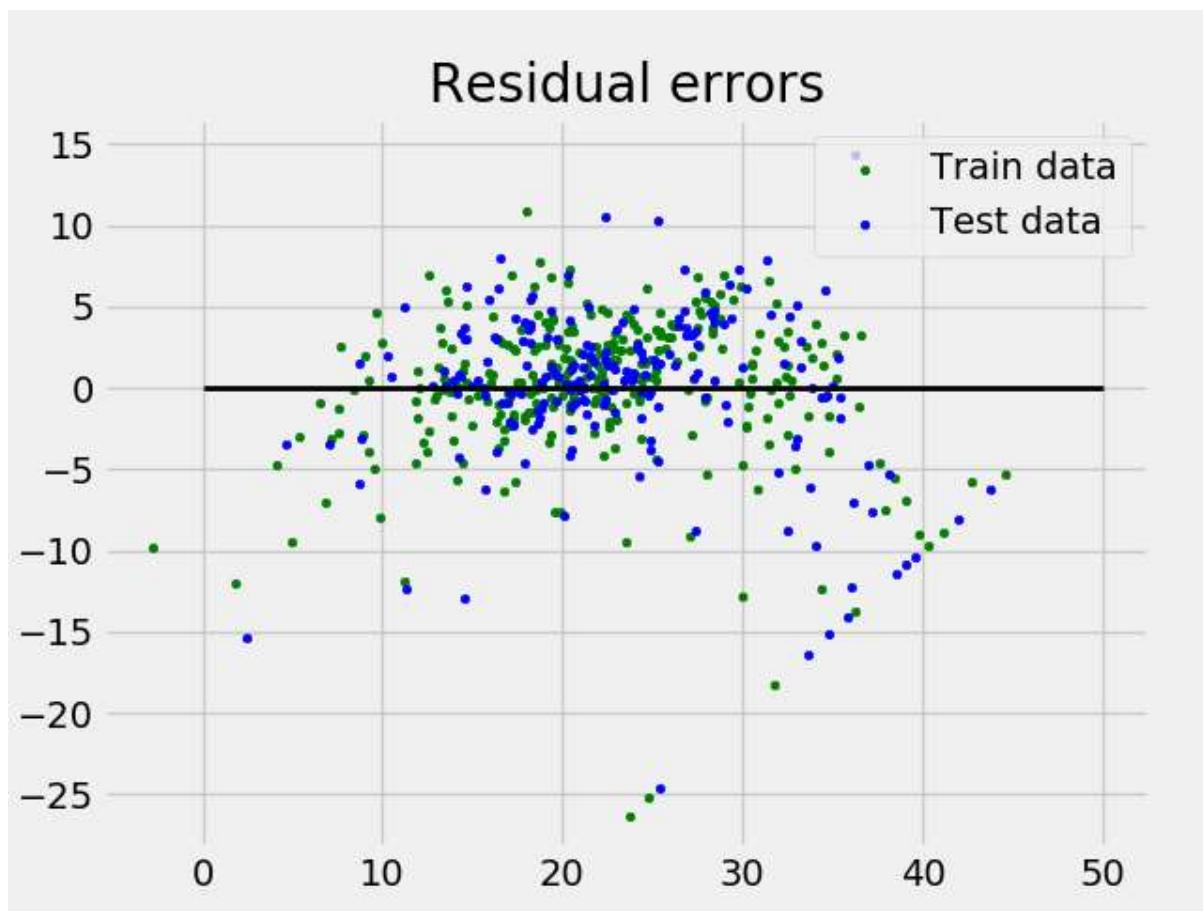
# plotting legend
plt.legend(loc='upper right')

# plot title
plt.title("Residual errors")

# method call for showing the plot
plt.show()
```

انتاج:

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)



مؤامرة الخطأ المتبقية للانحدار الخطى المتعدد

في المثال أعلاه، نحدد درجة الدقة باستخدام درجة التباين الموضحة . نحدد

$$\text{Var} / \{\text{ص} - \text{y}\} \text{ وأوضح فاريанс_سكور} = 1 - \text{فار}$$

هو التباين، مربع Var هو الناتج المستهدف المقابل (الصحيح)، و y ، هو الناتج المستهدف المقدر \hat{y} حيث الانحراف المعياري. أفضل نتيجة ممكنة هي 1.0 ، والقيمة الأقل هي الأسوأ.

الانحدار الخطى متعدد الحدود

\times الانحدار متعدد الحدود هو شكل من أشكال الانحدار الخطى حيث يتم تصميم العلاقة بين المتغير المستقل يناسب الانحدار متعدد الحدود العلاقة غير الخطية . n على أنها متعددة الحدود من الدرجة y والمتغير التابع x والذي يُشار إليه بـ ، y والمتوسط الشرطي المقابل لـ $E(y | x)$.

اختيار درجة للانحدار متعدد الحدود

إن اختيار درجة الانحدار متعدد الحدود هو عبارة عن مقايسة بين التحييز والتباين. التحييز هو ميل النموذج إلى التنبؤ باستمرار بنفس القيمة، بعض النظر عن القيمة الحقيقة للمتغير التابع. التباين هو ميل النموذج إلى تقديم تنبؤات مختلفة لنفس نقطة البيانات، اعتماداً على بيانات التدريب المحددة المستخدمة.

البيان ولكنها يمكن أن تزيد أيضًا من التحيز.

هناك عدد من الطرق لاختيار درجة الانحدار متعدد الحدود، مثل التحقق المتبادل واستخدام معايير المعلومات مثل معيار المعلومات الباييزية Akaike (AIC) أو معيار المعلومات الباييزية (BIC).

تنفيذ الانحدار متعدد الحدود باستخدام بايثون

تنفيذ الانحدار متعدد الحدود باستخدام بايثون

سنقوم هنا باستيراد جميع المكتبات اللازمة لتحليل البيانات ومهام التعلم الآلي، ثم نقوم بتحميل مجموعة البيانات ثم يقوم بإعداد البيانات للنموذج من خلال التعامل مع القيم "Position_Salaries.csv" باستخدام Pandas. ثم يقوم بـ تقسيم البيانات إلى مجموعات تدريب وختبار وتحديد الميزات المفقودة وترميز البيانات الفئوية. وأخيرًا، يقوم بـ تقسيم البيانات إلى مجموعات تدريب وختبار وتحديد الميزات الرقمية باستخدام StandardScaler.

Python3

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
import warnings

from sklearn.preprocessing import LabelEncoder
from sklearn.impute import KNNImputer
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import f1_score
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import cross_val_score

warnings.filterwarnings('ignore')

df = pd.read_csv('Position_Salaries.csv')

X = df.iloc[:,1:2].values
y = df.iloc[:,2].values
x
```

انتاج:

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

```
[ 3, 60000, 8],
[ 4, 80000, 5],
[ 5, 110000, 3],
[ 6, 150000, 7],
[ 7, 200000, 6],
[ 8, 300000, 9],
[ 9, 500000, 1],
[ 10, 1000000, 2]], dtype=int64)
```

يقوم الكود بإنشاء نموذج انحدار خطى ويلائمه مع البيانات المقدمة، مما ينشئ علاقة خطية بين المتغيرات المستقلة والتابعة.

Python3

```
from sklearn.linear_model import LinearRegression
lin_reg=LinearRegression()
lin_reg.fit(X,y)
```

ينفذ الكود انحداراً تربعياً ومكعبياً عن طريق إنشاء ميزات متعددة الحدود من البيانات الأصلية ونماذج الانحدار الخطى المناسبة لهذه الميزات. وهذا يتبع نمذجة العلاقات غير الخطية بين المتغيرات المستقلة والتابعة.

Python3

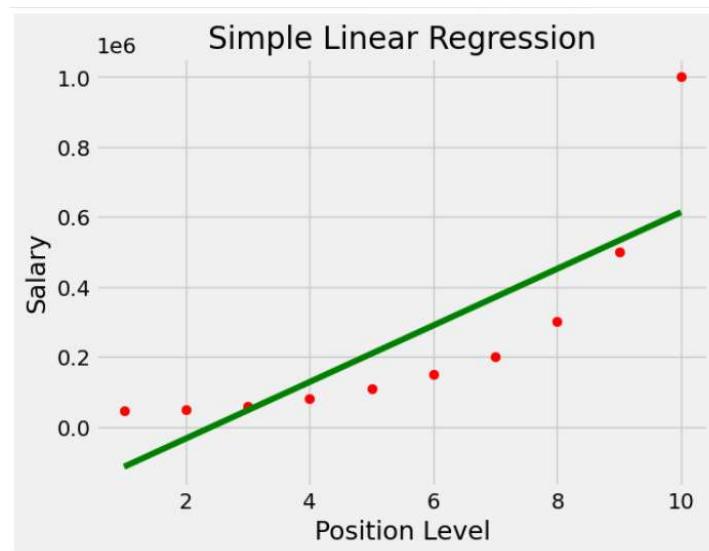
```
from sklearn.preprocessing import PolynomialFeatures
poly_reg2=PolynomialFeatures(degree=2)
X_poly=poly_reg2.fit_transform(X)
lin_reg_2=LinearRegression()
lin_reg_2.fit(X_poly,y)

poly_reg3=PolynomialFeatures(degree=3)
X_poly3=poly_reg3.fit_transform(X)
lin_reg_3=LinearRegression()
lin_reg_3.fit(X_poly3,y)
```

Python3

```
plt.scatter(X,y,color='red')
plt.plot(X,lin_reg.predict(X),color='green')
plt.title('Simple Linear Regression')
plt.xlabel('Position Level')
plt.ylabel('Salary')
plt.show()
```

:انتاج



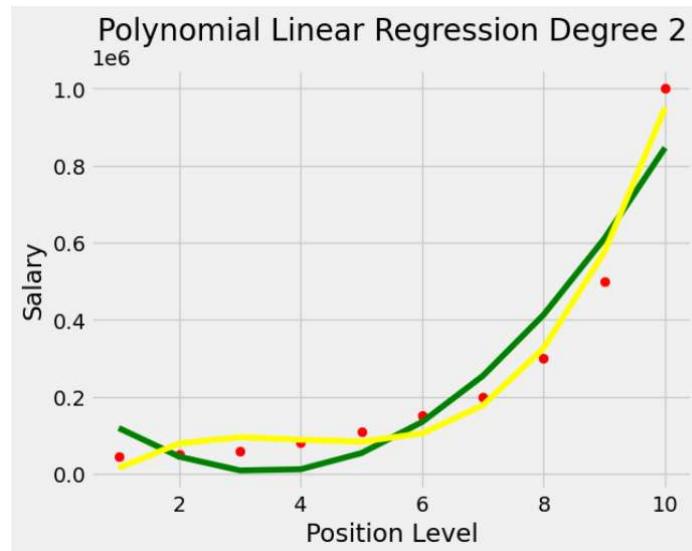
يقوم الكود بإنشاء مخطط مبعثر لنقاط البيانات، ويغطي خطوط الانحدار التربيعية والمكعبية المتوقعة. إنه يتصور بشكل فعال العلاقة غير الخطية بين مستوى المنصب والراتب ويقارن تناسبات نماذج الانحدار التربيعية والمكعبية.

Python3

```
plt.style.use('fivethirtyeight')
plt.scatter(X,y,color='red')
plt.plot(X,lin_reg_2.predict(poly_reg2.fit_transform(X)),color='green')
plt.plot(X,lin_reg_3.predict(poly_reg3.fit_transform(X)),color='yellow')
plt.title('Polynomial Linear Regression Degree 2')
plt.xlabel('Position Level')
plt.ylabel('Salary')
plt.show()
```

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

:انتاج



يتصور الكود بشكل فعال العلاقة بين مستوى المنصب والراتب باستخدام الانحدار المكعب ويولد خط تنبؤ مستمر لنطاق أوسع من مستويات المناصب.

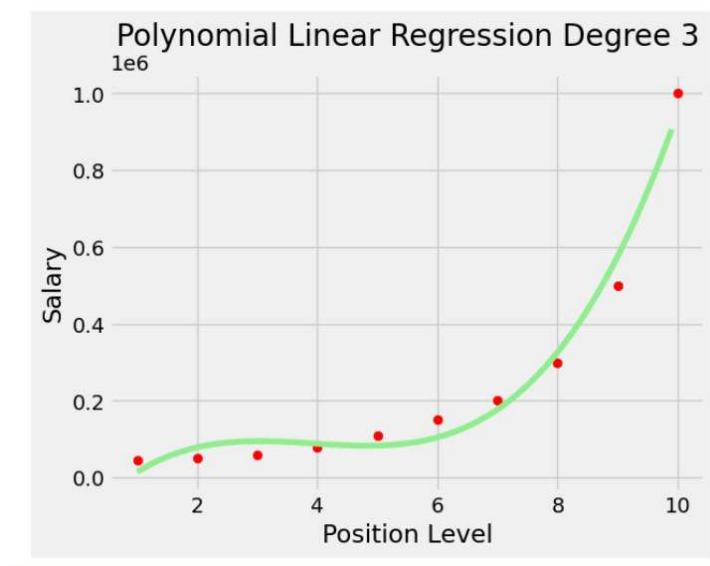
مشاريع التعلم الآلي | R | OpenCV | الباتدا | NumPy | دروس التعلم الآلي | دروس تحليل البيانات | بايثون - البرنامج التعليمي لتصور البيانات

```
plt.style.use('fivethirtyeight')
X_grid=np.arange(min(X),max(X),0.1) # This will give us a vector.We will have to convert it into a column vector
X_grid=X_grid.reshape((len(X_grid),1))
plt.scatter(X,y,color='red')
plt.plot(X_grid,lin_reg_3.predict(poly_reg3.fit_transform(X_grid)),color='lightgreen')

# plt.plot(X,lin_reg_3.predict(poly_reg3.fit_transform(X)),color='green')
plt.title('Polynomial Linear Regression Degree 3')
plt.xlabel('Position Level')
plt.ylabel('Salary')
plt.show()
```

:انتاج

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و [سياسة الخصوصية الخاصة بنا](#)



تطبيقات الانحدار الخطى

- خطوط الاتجاه:** يمثل خط الاتجاه التباين في البيانات الكمية مع مرور الوقت (مثل الناتج المحلي الإجمالي، وأسعار النفط، وما إلى ذلك). عادة ما تتبع هذه الاتجاهات علاقة خطية. وبالتالي، يمكن تطبيق الانحدار الخطى للتنبؤ بالقيم المستقبلية. ومع ذلك، فإن هذه الطريقة تعانى من نقص الصحة العلمية في الحالات التي يمكن أن تؤثر فيها التغيرات المحتملة الأخرى على البيانات.
- الاقتصاد:** الانحدار الخطى هو الأداة التجريبية السائدة في الاقتصاد. على سبيل المثال، يتم استخدامه للتنبؤ بالإنفاق الاستهلاكى، والإنفاق الاستثماري الثابت، واستثمار المخزون، ومشتريات صادرات الدولة، والإنفاق على الواردات، والطلب على الاحتياط بالأصول السائلة، والطلب على العمالة، وعرض العمالة.
- التمويل:** يستخدم نموذج أصول سعر رأس المال الانحدار الخطى لتحليل وقياس المخاطر المنهجية للاستثمار.
- علم الأحياء:** يستخدم الانحدار الخطى لنموذج العلاقات السببية بين المعلمات في النظم البيولوجية.

مزایا الانحدار الخطى

- سهولة التفسير:** تمثل معاملات نموذج الانحدار الخطى التغير في المتغير التابع لتغير وحدة واحدة في المتغير المستقل، مما يجعل من السهل فهم العلاقة بين المتغيرات.
- قوى للمترفة:** الانحدار الخطى قوى نسبياً لقيم المترفة مما يعني أنه أقل تأثراً بالقيم المترفة للمتغير المستقل مقارنة بالطرق الإحصائية الأخرى.
- يمكنه التعامل مع العلاقات الخطية وغير الخطية:** يمكن استخدام الانحدار الخطى لنموذج العلاقات الخطية وغير الخطية بين المتغيرات. وذلك لأن المتغير المستقل يمكن تحويله قبل استخدامه في النموذج.
- لا حاجة إلى قياس الميزات أو تحويلها:** على عكس بعض خوارزميات التعلم الآلي، لا يتطلب الانحدار الخطى تغيير حجم الميزات أو تحويلها. يمكن أن يكون هذا ميزة كبيرة، خاصة عند التعامل مع مجموعات البيانات الكبيرة.

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت [سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

- يفرض الخطية:** يفترض الانحدار الخطي أن العلاقة بين المتغير المستقل والمتغير التابع خطية. قد لا يكون هذا الافتراض صالحًا لجميع مجموعات البيانات. في الحالات التي تكون فيها العلاقة غير خطية، قد لا يكون الانحدار الخطي خيارًا جيداً.
- حساس للتعددية الخطية:** الانحدار الخطي حساس للتعددية الخطية. يحدث هذا عندما يكون هناك ارتباط كبير بين المتغيرات المستقلة. يمكن أن يجعل العلاقة الخطية المتعددة من الصعب تفسير معاملات النموذج ويمكن أن يؤدي إلى التجهيز الزائد.
- قد لا يكون مناسباً للعلاقات شديدة التعقيد:** قد لا يكون الانحدار الخطي مناسباً لنمذجة العلاقات شديدة التعقيد بين المتغيرات. على سبيل المثال، قد لا يكون قادراً على نمذجة العلاقات التي تتضمن التفاعلات بين المتغيرات المستقلة.
- غير مناسب لمهام التصنيف:** الانحدار الخطي هو خوارزمية انحدار وغير مناسب لمهام التصنيف التي تتضمن التنبؤ بمتغير فوري بدلاً من متغير مستمر.

الأسئلة المتداولة (الأسئلة الشائعة)

1. كيفية استخدام الانحدار الخطي لإجراء التنبؤات؟

بمجرد تدريب نموذج الانحدار الخطي، يمكن استخدامه للتنبؤ بنقاط البيانات الجديدة. طريقة تسمى `التنبؤ()` والتي يمكن `LinearRegression` من `scikit-learn` توفر فئة. استخدامها لإجراء التنبؤات.

2. ما هو الانحدار الخطي؟

الانحدار الخطي هو خوارزمية تعلم آلي خاضعة للإشراف تستخدم للتنبؤ بالمخرجات الرقمية المستمرة. ويفترض أن العلاقة بين المتغيرات المستقلة (الميزات) والمتغير التابع (الهدف) خطية، مما يعني أنه يمكن حساب القيمة المتوقعة للهدف كمجموع خطية من الميزات.

3. كيفية تنفيذ الانحدار الخطي في بايثون؟

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت [سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

الأكثر شيوعاً لمهام التعلم الآلى، وتتوفر تنفيئاً بسيطاً وفعالاً للانحدار الخطى.

ما هي بعض تطبيقات الانحدار الخطى؟ 4.

الانحدار الخطى هو خوارزمية متعددة الاستخدامات يمكن استخدامها لمجموعة واسعة من التطبيقات، بما في ذلك التمويل والرعاية الصحية والتسويق. تتضمن بعض الأمثلة المحددة ما يلى:

- التنبؤ بأسعار المنازل
- التنبؤ بأسعار الأسهم
- تشخيص الحالات الطبية
- التنبؤ بحركة العملاء

كيف يتم تنفيذ الانحدار الخطى في 5.

توفر scikit-learn باستخدامر فئة `LinearRegression`. يتم تنفيذ الانحدار الخطى في هذه الفئة طرقة لملاءمة نموذج الانحدار الخطى لمجموعة بيانات التدريب والتنبؤ بالقيمة المستهدفة لنقاط البيانات الجديدة.

سواء كنت تستعد لمقابلة العمل الأولى أو تهدف إلى تحسين مهاراتك في هذا المشهد التكنولوجى المتتطور هي مفتاحك للنجاح. نحن نقدم محتوى عالي الجودة بأسعار معقولة، [GeeksforGeeks](#) باستمرار، فإن دورات وكلها موجهة نحو تسريع نموك بطريقة محددة زمنياً. انضم إلى الملابين الذين قمنا بتمكينهم بالفعل، ونحن هنا لنفعل نفس الشيء من أجلك. لا تفوت - [التحقق من ذلك الآن](#)

آخر تحديث: 07 ديسمبر 2023

87

سابق

التالى

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت [سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

قراءات مماثلة

مل | الانحدار الخطى مقابل الانحدار اللوجستى

تنفيذ الانحدار الخطى من الصفر باستخدام بايثون

تنفيذ الانحدار الخطى المرجح محليا

مل | الانحدار الخطى المتعدد باستخدام بايثون

الانحدار الخطى sklearn بايثون | الانحدار الخطى باستخدام

حل الانحدار الخطى في بايثون

الانحدار الخطى المرجح محليا باستخدام بايثون

بايثون | تنفيذ الانحدار متعدد الحدود

تنفيذ انحدار ريدج من الصفر باستخدام بايثون

تنفيذ انحدار لاسو من الصفر باستخدام بايثون

دورس كاملة

الدليل الكامل - OpenAI Python API

دورس رؤية الكمبيوتر

علوم الكمبيوتر والبرمجة للأطفال

الباندا الذكاء الاصطناعي: مكتبة بايثون للذكاء الاصطناعي التوليدى

أهم مشاريع الرؤية الحاسوبية (2023)



GeeksforGeeks

التعلم الالى ، بيثون : علامات المادة

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

معلومات إضافية



Company

- [About Us](#)
- [Legal](#)
- [Careers](#)
- [In Media](#)
- [Contact Us](#)
- [Advertise with us](#)
- [GFG Corporate Solution](#)
- [Placement Training Program](#)
- [Apply for Mentor](#)
- [Job-A-Thon Hiring Challenge](#)
- [Hack-A-Thon](#)
- [GfG Weekly Contest](#)
- [Offline Classes \(Delhi/NCR\)](#)
- [DSA in JAVA/C++](#)
- [Master System Design](#)
- [Master CP](#)
- [GeeksforGeeks Videos](#)

Explore

Languages

[Python](#)

DSA

[Data Structures](#)

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و [سياسة الخصوصية الخاصة بنا](#)

C++	DSA for Beginners
PHP	Basic DSA Problems
GoLang	DSA Roadmap
SQL	Top 100 DSA Interview Problems
R Language	DSA Roadmap by Sandeep Jain
Android Tutorial	All Cheat Sheets

Data Science & ML

Data Science With Python
 Data Science For Beginner
 Machine Learning Tutorial
 ML Maths
 Data Visualisation Tutorial
 Pandas Tutorial
 NumPy Tutorial
 NLP Tutorial
 Deep Learning Tutorial

HTML & CSS

HTML
 CSS
 Bootstrap
 Tailwind CSS
 SASS
 LESS
 Web Design

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و [سياسة الخصوصية الخاصة بنا](#)

Python

- Python Programming Examples
- Django Tutorial
- Python Projects
- Python Tkinter
- Web Scraping
- OpenCV Python Tutorial
- Python Interview Question

Computer Science

- GATE CS Notes
- Operating Systems
- Computer Network
- Database Management System
- Software Engineering
- Digital Logic Design
- Engineering Maths

DevOps

- Git
- AWS
- Docker
- Kubernetes
- Azure
- GCP
- DevOps Roadmap

Competitive Programming

- Top DS or Algo for CP
- Top 50 Tree
- Top 50 Graph
- Top 50 Array
- Top 50 String
- Top 50 DP
- Top 15 Websites for CP

System Design

- What is System Design
- Monolithic and Distributed SD
- High Level Design or HLD
- Low Level Design or LLD
- Crack System Design Round
- System Design Interview Questions
- Grokking Modern System Design

JavaScript

- TypeScript
- ReactJS
- NextJS
- AngularJS
- NodeJS
- Express.js
- Lodash
- Web Browser

NCERT Solutions

- Class 12
- Class 11
- Class 10
- Class 9

School Subjects

- Mathematics
- Physics
- Chemistry
- Biology

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)

Complete Study Material

English Grammar

Commerce

- Accountancy
- Business Studies
- Indian Economics
- Macroeconomics
- Microeconomics
- Statistics for Economics

Management & Finance

- Management
- HR Management
- Income Tax
- Finance
- Economics

UPSC Study Material

- Polity Notes
- Geography Notes
- History Notes
- Science and Technology Notes
- Economy Notes
- Ethics Notes
- Previous Year Papers

SSC/ BANKING

- SSC CGL Syllabus
- SBI PO Syllabus
- SBI Clerk Syllabus
- IBPS PO Syllabus
- IBPS Clerk Syllabus
- SSC CGL Practice Papers

Colleges

- Indian Colleges Admission & Campus Experiences
- Top Engineering Colleges
- Top BCA Colleges
- Top MBA Colleges
- Top Architecture College
- Choose College For Graduation

Companies

- IT Companies
- Software Development Companies
- Artificial Intelligence(AI) Companies
- CyberSecurity Companies
- Service Based Companies
- Product Based Companies
- PSUs for CS Engineers

Preparation Corner

- Company Wise Preparation
- Preparation for SDE
- Experienced Interviews
- Internship Interviews

Exams

- JEE Mains
- JEE Advanced
- GATE CS
- NEET

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك توافق بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و [سياسة الخصوصية الخاصة بنا](#)

Puzzles

More Tutorials

Software Development
Software Testing
Product Management
SAP
SEO
Linux
Excel

Write & Earn

Write an Article
Improve an Article
Pick Topics to Write
Share your Experiences
Internships

@GeeksforGeeks, Sanchhaya Education Private Limited, All rights reserved

نحن نستخدم ملفات تعريف الارتباط لضمان حصولك على أفضل تجربة تصفح على موقعنا. باستخدام موقعنا، فإنك تقر بأنك قد قرأت وفهمت
[سياسة ملفات تعريف الارتباط](#) و[سياسة الخصوصية الخاصة بنا](#)