جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل / كلية التربية للعلوم الصرفة
قسم الرياضيات

الانحدار الخطي والانحدار اللوجستي

الطالب
كرار طعمه عبد عطية
بإشراف
م.م ايناس

بحث مقدم الى مجلس كلية التربية للعلوم الصرفة في جامعة بابل
وهو كجزء من متطلبات نيل شهادة البكالوريوس

١٤٤٥هـ ٢٠٢٤م

الآية

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

((ومن يؤت الحكمة فقد اؤتي خيرا كثيرا))

صدق الله العلي العظيم

البقرة ،٢٦٩

الإهداء

اهدي عملي هذا إلى فخر الكائنات وأفضل الانبياء وال الطاهرين محمد (صلى الله عليه واله وسلم)

الى كل من كلله الله بالهيبة والوفاء إلى من احمل اسمه بكل افتخار والدي رحمه الله

الى من كان دعائها سر نجاحي وحنانها بلسم جراحي والدتي

الى كل هؤلاء اهدي ثمرة جهدي المتواضع مع الحب والتقدير

الشكر والتقدير

الحمد الله رب العالمين الشافي، الكافي، المعافي،

والصلاة والسلام على سيد الخلق محمد بن عبد الله (صلى الله عليه واله وسلم)

يسر الباحث وقد شارف على الانتهاء من بحثه أن يتقدم

بجزيل الشكر وأعظم الامتنان إلى الدكتور ايناس المشرفة على البحث بجهودها ونصائحها

وتوجيهاتها وداعيا الباري أن يحفظها ويمن عليها بالصحة والعافية

المحتويات

الآية ١

الإهداء ٢
الشكر والتقدير ٣
مقدمة: ٤
الفصل الاول ٨
الانحدار الخطي والانحدار اللوجستي ٨
(1.1) الانحدار الخطي ٨
(1.1.1)الانحدار الخطي البسيط ٩
(2.1.1) الانحدار الخطي المتعدد ١١
(2.1) مزايا الانحدار الخطي: ١٣
(3.1) الانحدار اللوجستي ١٤
(1.3.1) الانحدار اللوجستي ثنائي الاستجابة ١٤
(2.3.1) الانحدار اللوجستي المتعدد ١٦
(41.) مزايا الانحدار اللوجستي: ١٧
الفصل الثاني ١٨
مقارنة بين الانحدار الخطي واللوجستي ١٨
(12.) مقارنة بين الانحدار الخطي واللوجستي في الهدف والاستخدامات: ١٨
(2.2) المقارنة بين الأداء الإحصائي للانحدار الخطي والانحدار اللوجستي: ٢٠
(32.) الاستخدام العملي لكل من الانحدار الخطي واللوجستي: ٢٠
الفصل الثالث: ٢٣
كيف يعمل الانحدار الخطي واللوجستي في التعلم الآلي مع تطبيق عملي في Python ٢٣
(2.3) الانحدار الخطي في تعلم الالة ٢٤
(3.3) الانحدار اللوجستي في تعلم الالة ٢٥
الفصل الرابع ٢٩
الاستنتاجات ٢٩
المصادر ٣٠

مقدمة:

يعرف الانحدار على انه أسلوب احصائي لدراسة ونمذجة العلاقة بين المتغيرات (Montgomery, peck & vining, 2006) حيث يتم دراسة انحدار او اعتماد متغير واحد

يسمى المتغير التابع dependent variable على متغير واحد او أكثر تسمى بالمتغيرات المستقلة independent variable وبناء نموذج رياضي يصف علاقة الارتباط بين المتغير التابع والمتغيرات المستقلة.

ان أغراض تحليل الانحدار أربعة:

١. وصف البيانات data description. أي وصف شكل علاقة الارتباط بين المتغير التابع والمتغيرات المستقلة.
٢. تقدير المعلمات parameters estimation. أي يتم استخدام تحليل الانحدار لتقدير معلمات بعض النماذج الرياضية المعروفة التي تربط بين متغير تابع ومتغيرات مستقلة معينة.
٣. التقدير والتنبؤ prediction and estimation. أي التقدير والتنبؤ بقيم المتغير التابع بمعلومية قيم المتغيرات المستقلة.
٤. التحكم control. أي التحكم بقيم المتغير التابع وفقا لقيم محدودة من المتغيرات المستقلة، وهذا الهدف الأخير يتم تحقيقه في حالة العلاقة السببية بين المتغير التابع والمتغيرات المستقلة. (الاحصاء التطبيقي سعد)

ينقسم الانحدار الى أنواع عديدة تناسب ظروفا مختلفة متعلقة بنوع وعدد المتغيرات المستقلة والتابعة، وشكل وتوزيع المتغيرات وارتباطها ببعضها وحجم العينة. **(مقال من QUESTION PRO)** وهي:

١. الانحدار الخطي
٢. الانحدار اللوجستي
٣. الانحدار متعدد الحدود
٤. الانحدار التدريجي
٥. انحدار ريدج
٦. انحدار اللاسو
٧. الانحدار الصافي المرن

وسنكتفي بالحديث على أكثر أساليب الانحدار استخداما وأكثرها مألوفيه، **وهما الانحدار الخطي والانحدار اللوجستي**

هدف البحث

في الانحدار الخطي واللوجستي يتمثل في فهم العلاقة بين المتغيرات وتحليل البيانات بطريقة إحصائية لتحقيق أهداف معينة. من بين أهداف البحث الرئيسية لكل منهما:

هدف البحث في الانحدار الخطي:

١ .تحليل العلاقة الخطية: الهدف الرئيسي هو فهم طبيعة العلاقة بين متغير تابع واحد أو أكثر ومتغيرات مستقلة.

٢ .توقع القيم: يُستخدم الانحدار الخطي لتوقع قيم المتغير التابع بناءً على قيم المتغيرات المستقلة.

٣ .تحديد العوامل المؤثرة: يُمكن استخدام الانحدار الخطي لتحديد العوامل المستقلة التي تؤثر بشكل كبير على المتغير التابع.

٤ .فهم التأثيرات: يُمكن استخدام الانحدار الخطي لتقدير مقدار التأثير أو الزيادة في المتغير التابع نتيجة لتغير في المتغيرات المستقلة.

هدف البحث في الانحدار اللوجستي:

١ .تحليل العلاقة الاحتمالية: الهدف الرئيسي هو فهم العلاقة بين متغير تابع ثنائي القيمة ومتغيرات مستقلة.

٢ .توقع الاحتمالات: يُستخدم الانحدار اللوجستي لتوقع احتمالات الحدوث أو عدم الحدوث للحدث الثنائي بناءً على قيم المتغيرات المستقلة.

٣ .تحديد العوامل المؤثرة: يُمكن استخدام الانحدار اللوجستي لتحديد العوامل المستقلة التي تؤثر بشكل كبير على احتمالات الحدوث للحدث الثنائي.

٤ .تقدير الأثر: يُمكن استخدام الانحدار اللوجستي لتقدير مدى تأثير كل عامل مستقل على احتمالات الحدوث للحدث الثنائي.

الفصل الاول

الانحدار الخطي والانحدار اللوجستي

الفصل الأول

الانحدار الخطي والانحدار اللوجستي

(1.1) الانحدار الخطي

هو تقنية إحصائية تُستخدم لفهم العلاقة بين متغير تابع ومتغير أو أكثر من المتغيرات المستقلة. يُمثل الهدف الرئيسي للانحدار الخطي تقدير العلاقة الخطية بين المتغيرات المستقلة والتابعة، وذلك من خلال إيجاد نموذج يمثل هذه العلاقة بشكل مبسط ومفهوم.

ينقسم الانحدار الخطي حسب عدد المتغيرات المستقلة الى قسمين:

(١.١.١) الانحدار الخطي البسيط

في هذا النوع يكون لدينا متغيران أحدهما تابع والأخر مستقل والعلاقة بينهما خطية. ويستخدم للتنبؤ بقيم المتغير التابع بدلا من المتغير المستقل فمثلا قد يرغب باحث في تحديد شكل العلاقة بين وزن الطفل (متغير تابع) وعمره (متغير مستقل)، العلاقة بين حجم سلعة ما (متغير تابع) وحجم مصروفات الدعاية (متغسر مستقل)، العلاقة بين مستوى الأداء الوظيفي (متغير تابع) والمؤهل الأكاديمي (متغير مستقل) ...الخ. (كتاب تحليل الانحدار الخطي)

ونرمز للمتغير التابع بالحرف y وللمتغير المستقل بالحرف x. ويسمى النموذج في هذه الحالة بنموذج انحدار y على x.

وان البيانات التي تستخدم في بناء نموذج الانحدار البسيط تحتوي على (n) مشاهدة حول المتغير التابع مع المتغير المستقل (جدول ١.١)

جدول (١.١): بيانات نموذج الانحدار الخطي البسيط

رقم المشاهدة	قيم المتغير التابع Y (y_i)	قيم المتغير المستقل X (x_i)
1	y_1	x_1
2	y_2	x_2
3	y_3	x_3
4	y_4	x_4
.	.	.
.	.	.
.	.	.
n	y_n	x_n

ان ابسط علاقة دالية تربط بين المتغيرين x وy يمكن التعبير عنها على النحو الاتي:

$$Y = f(x)$$

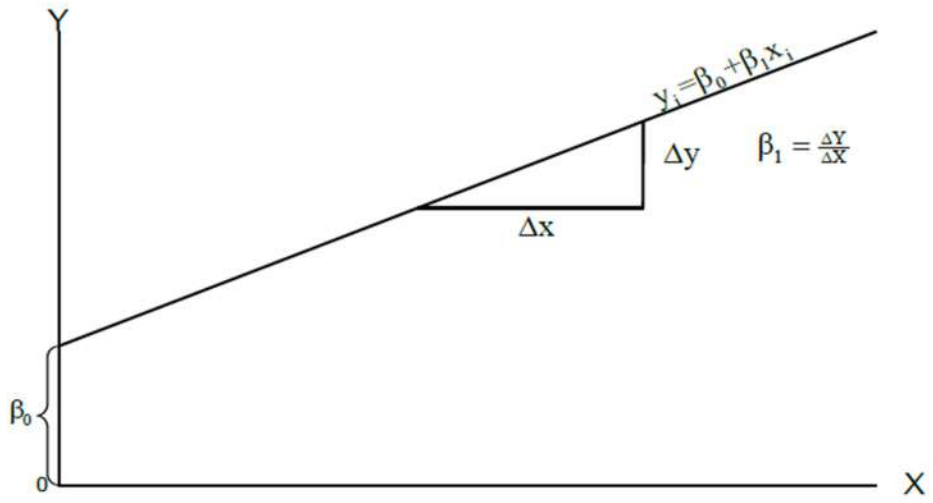
حيث ان $f(x)$ يرمز الى ان y دالة (function) ل x، أي ان قيمة Y تتغير تبعا لتغير قيمة x ولكن هذه المعادلة لا تحدد شكل العلاقة بين المتغيرين، فقد تكون الدالة على صورة خطية او منحنى اخر، ويعتمد تحديد صيغة الدالة على الباحث وذلك بوضع افتراضات حول هذه العلاقة.

فمثلا قد يفترض باحث ان هناك علاقة خطية بين الادخال وسعر الفائدة وبهذا التقريب لشكل الدالة يمكن ان تأخذ المعادلة صيغة الدالة الخطية التالية: (كتاب تحليل الانحدار الخطي)

$$Y = \beta_0 + \beta_1 X$$

حيث ان Y المتغير التابع و X المتغير المستقل وβ_0 و β_1 معالم مجهولة تعرف بمعاملات الانحدار.

ويعرف β_0 بالمعامل الثابت او المقطع الصادي و β_1 بمعامل الانحدار او ميل الدالة الخطية، وتسمى هذه المعادلة بمعادلة الانحدار الخطي للمجتمع.



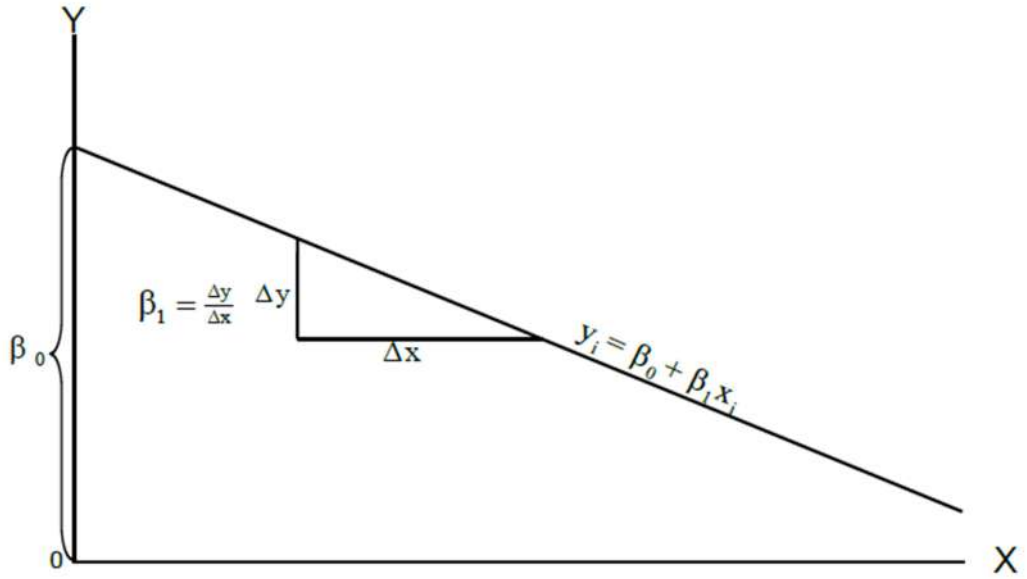
الشكل (1.1): نموذج انحدار خطي بسيط حالة ميل موجب $\beta_1 > 0$

الشكل (2.1): نموذج انحدار خطي بسيط حالة ميل سالب $\beta_1 < 0$

(٢.١.١) الانحدار الخطي المتعدد

في هذا النوع متغير مستقل واحد على المتغير التابع. الا انه نادرا ما نجد متغيرا واحدا يفسر جزئا كبيرا من التغير او التباين في المتغير التابع. لذلك نجد ان نموذج الانحدار الخطي المتعدد الذي يقيس اثر اكثر من متغير واحد على المتغير التابع هو الاوسع استخداما.

ويعتبر نموذج الانحدار الخطي المتعدد تعميما للمفاهيم والأساليب والصيغ المستخدمة في نموذج



الانحدار الخطي البسيط، ويتمثل الفرق الوحيد بين نموذجي الانحدار الخطي البسيط والانحدار الخطي المتعدد في ان الأول يضم متغير مستقل واحد في حين يضم الثاني متغيرين مستقلين او اكثر مكن المتغيرات المستقلة. والجدول (2.1) يوضح بيانات نموذج الانحدار الخطي المتعدد المتكونة من (n) من المشاهدات للمتغير التابع مع (p) من المتغيرات المستقلة.

الجدول (2.1): بيانات نموذج الانحدار الخطي المتعدد

رقم المشاهدة	المتغير التابع	المتغيرات المستقلة			
	y_i	x_{1i}	x_{2i}	. . .	x_{pi}
1	y_1	x_{11}	x_{21}	. . .	x_{p1}
2	y_2	x_{12}	x_{22}	. . .	x_{p2}
3	y_3	x_{13}	x_{23}	. . .	x_{p3}
.	.	.	.	.	.
n	y_n	x_{1n}	x_{2n}	. . .	x_{pn}

معادلة نموذج الانحدار الخطي المتعدد الذي يضم عدد (p) متغير مستقل بالصيغة التالية:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i \qquad i = 1,2,\ldots,N$$

$$= \beta_0 + \sum_{j=1}^{p} \beta_j x_{ji} + \varepsilon_i \qquad \textbf{معادلة (1.1)}$$

حيث ان: Y المتغير التابع.

و $\beta_0, \beta_1, \beta_2, \ldots\ldots, \beta_p$ معالم النموذج المجهولة المراد تقديرها وتسمى أيضا بمعاملات الانحدار الجزئية. و $X_0, X_1, X_2, \ldots\ldots, X_p$ المتغيرات المستقلة. و N عدد المشاهدات.

والمعادلة أعلاه هي اختصار لعدد N من المعادلات الانية التالية:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{21} + \ldots + \beta_p x_{p1} + \varepsilon_1$$
$$y_2 = \beta_0 + \beta_1 x_{12} + \beta_2 x_{22} + \ldots + \beta_p x_{p2} + \varepsilon_2$$
$$y_3 = \beta_0 + \beta_1 x_{13} + \beta_2 x_{23} + \ldots + \beta_p x_{p3} + \varepsilon_3$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$y_N = \beta_0 + \beta_1 x_{1N} + \beta_2 x_{2N} + \ldots + \beta_p x_{pN} + \varepsilon_N \qquad \textbf{معادلة (1.2)}$$

وباستخدام صيغ المصفوفات يمكن كتابتها على النحو التالي: (كتاب تحليل الانحدار الخطي)

وباستخدام رموز المصفوفات يمكن كتابة المعادلات كما يلي:

$$Y = X\beta + \varepsilon$$

(٢.١) مزايا الانحدار الخطي:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & . & . & . & x_{p1} \\ 1 & x_{12} & x_{22} & . & . & . & x_{p2} \\ 1 & x_{13} & x_{23} & . & . & . & x_{p3} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ 1 & x_{1N} & x_{2N} & & & & x_{pN} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_N \end{pmatrix}$$

معادلة (1.3)

١ .سهولة التفسير: يوفر الانحدار الخطي نموذجًا بسيطًا وسهل الفهم للعلاقة بين المتغيرات، مما يجعله مناسبًا للتفسير من قبل غير المتخصصين.

٢ .تنبؤ القيم: يُمكن استخدام الانحدار الخطي لتنبؤ القيم المستقبلية بناءً على العلاقة السابقة بين المتغيرات، مما يساعد في اتخاذ القرارات الاستراتيجية.

٣ .التحليل الكمي: يوفر الانحدار الخطي تحليلًا كميًّا للعلاقة بين المتغيرات، مما يسمح بتقدير قوة العلاقة واختبار فروض البحث.

٤ .التوافق مع البرامج الإحصائية: يتوافق الانحدار الخطي مع معظم برامج التحليل الإحصائي، مما يجعل عملية التحليل والتفسير أسهل للباحثين.

(3.1) الانحدار اللوجستي

الانحدار اللوجستي هو أحد الطرق الإحصائية المستخدمة لتحليل العلاقة بين متغير تابع ثنائي القيمة ومجموعة من المتغيرات المستقلة. يهدف الانحدار اللوجستي إلى تحديد الاحتمالات أو الاحتمالات المعدلة لحدوث حدث معين بناءً على المتغيرات المستقلة.

يستخدم الانحدار اللوجستي في العديد من المجالات مثل علوم الحياة والطب والعلوم الاجتماعية والاقتصاد والتسويق وغيرها. تُعد الدالة اللوجستية المستخدمة في الانحدار اللوجستي أساسية لتحويل نتائج الانحدار إلى نطاق بين ٠ و١، وهذا يمكن استخدامها لتوفير تقديرات الاحتمالات.

يتم تحقيق الانحدار اللوجستي عن طريق تطبيق تقنيات الحد الأدنى للمربعات أو التقنيات المبنية على الاحتمالات مثل الانحدار اللوجستي المحدد أو الانحدار اللوجستي العشوائي.

يعد الانحدار اللوجستي أداة قوية في تحليل البيانات واتخاذ القرارات التنبؤية. يمكن أن يساعد في فهم العوامل المؤثرة في حدوث حدث معين وتقدير الاحتمالات المرتبطة بهذا الحدث بناءً على المتغيرات المستقلة المتاحة. (رسالة دكتوراه عادل بابطين الانحدار اللوجستي)

ينقسم الانحدار اللوجستي الى نموذجين:

(1.3.1) الانحدار اللوجستي ثنائي الاستجابة

يعرف نموذج الانحدار اللوجستي الثنائي على أنه أحد نماذج الانحدار اللاخطية حيث يبنى على فروض أساسية لذلك فأن متغير الاستجابة الذي نهتم بدراسته يتبع توزيع برنولي باحتمال نجاح π_i عندما (y=1) واحتمال فشل $1 - \pi_i$ عندما (y=0) بذلك تكون صيغة دالة الكثافة الاحتمالية تكتب كالاتي:

معادلة (1.4) $$P(y_i \backslash X_i) = [\pi(X_i)]^{y_i} [1 - \pi(X_i)]^{1-y}$$

وعند تعويض قيم $y_i = 0,1$ سيكون p(y) فان:

$$P(Y) = \begin{cases} \pi_i & \text{عند حدوث الاستجابة } Y = 1 \text{ عندما} \\ 1 - \pi_i & \text{عند عدم حدوث الاستجابة } Y = 0 \text{ عندما} \end{cases}$$

حيث:

π_i: يمثل احتمال النجاح

$1 - \pi_i$: يمثل احتمال الفشل

وعندما $y_i = 1$ نحصل على:

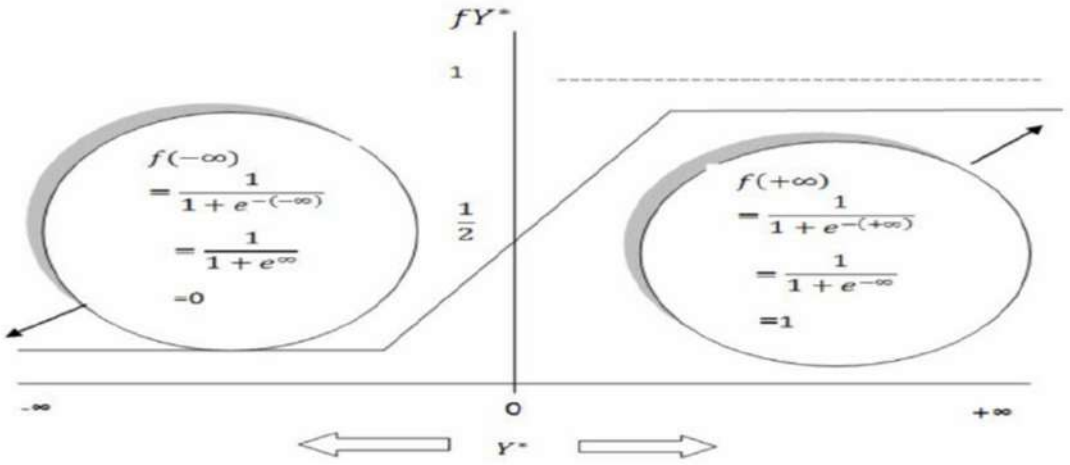
معادلة (1.5) $$\pi i = \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

وعندما $y_i = 0$ نحصل على:

$$1 - \pi i = 1 - \frac{e^{B_0 + B_1 X_1}}{1 + e^{B_0 + B_1 X_1}}$$

معادلة (1.6) $$1 - \pi i = \frac{1}{1 + e^{B_0 + B_1 X_1}}$$

ويمكن عرض الدالة اللوجستية بالشكل (3.1):



الشكل (3.1) الدالة اللوجستية

ومن الشكل أعلاه يتضح بان الدالة اللوجستية $f(y^*)$ هي دالة احتمالية تكون محصورة ضمن القيم (0,1)

(2.3.1) الانحدار اللوجستي المتعدد

يعتبر أنموذج الانحدار اللوجستي متعدد الاستجابة من النماذج الاحصائية ذات الأهمية الكبيرة في تحليل البيانات المصنفة، حيث يستخدم بصورة عامة في حالة إذا كان متغير الاستجابة يعود لمتغيرات التي تكون من النوع الاسمي او الرتبي والمكونة من تصنيفين أَو مستويين ف أكثر.

ويعد أنموذج الانحدار اللوجستي المتعدد الاستجابة امتداداً بسيطا لأنموذج الانحدار اللوجستي الثنائي الاستجابة. ويعتمد أنموذج الانحدار اللوجستي متعدد الاستجابة بصورة اساسية على التوزيع المتعدد الحدود.

$$P_r(r_{i1}, r_{i2}, \ldots\ldots, r_{ij}) = \binom{Ni}{r_{ij}, \ldots\ldots, r_{iJ}} \pi_{i1}^{r_{i1}} \ldots \pi_{ij}^{r_{ij}}$$

(٤١.) مزايا الانحدار اللوجستي:

١ .توفير احتماليات الحدوث: يوفر الانحدار اللوجستي احتماليات الحدوث للأحداث الثنائية، مما يساعد في تقدير فرص الحدوث وفهم العوامل المؤثرة فيها.

٢ .توافق مع البيانات الثنائية: يعتبر الانحدار اللوجستي ملائمًا لتحليل البيانات الثنائية أو الثنائية المتعددة، مثل النجاح والفشل، أو النعم واللا.

٣ .تحليل العوامل المؤثرة: يساعد الانحدار اللوجستي في تحليل العوامل المؤثرة في الظواهر الثنائية، مما يسمح بتحديد العوامل الرئيسية التي تؤثر في حدوث الحدث المرغوب أو غير المرغوب فيه.

الفصل الثاني
مقارنة بين الانحدار الخطي واللوجستي

الفصل الثاني

مقارنة بين الانحدار الخطي واللوجستي

(١٢.) مقارنة بين الانحدار الخطي واللوجستي في الهدف والاستخدامات:

١ .الهدف:

- الانحدار الخطي: الهدف الرئيسي من الانحدار الخطي هو تحليل العلاقة بين متغير تابع ومتغيرات مستقلة لتنبؤ القيم المستقبلية أو فهم العلاقة بينها.

- الانحدار اللوجستي: الهدف الرئيسي من الانحدار اللوجستي هو تحليل العلاقة بين متغير تابع ثنائي القيمة ومتغيرات مستقلة لتوقع احتمالات الحدوث للحدث الثنائي.

٢ .التوزيع:

- الانحدار الخطي: يفترض الانحدار الخطي توزيعًا طبيعيًا للبيانات، ويعتمد على تقدير معاملات يُمثل العلاقة الخطية بين المتغيرات.

- الانحدار اللوجستي: يفترض الانحدار اللوجستي توزيعًا للاحتمالات بدلاً من القيم المستقلة، ويستخدم اللوغاريتم الطبيعي لتمثيل العلاقة بين المتغيرات.

٣ .الاستخدامات:

- الانحدار الخطي: يستخدم الانحدار الخطي في تحليل البيانات المستمرة والتنبؤ بالقيم المستقبلية لها، مثل توقع المبيعات أو تحليل العوامل المؤثرة في درجات الطلاب.

- الانحدار اللوجستي: يستخدم الانحدار اللوجستي في تحليل البيانات الثنائية وتوقع احتمالات الحدوث للحدث الثنائي، مثل تحليل عوامل الخطر للإصابة بمرض معين أو توقع احتمالات النجاح في اختبار معين.

٤ .التفسير:

- الانحدار الخطي: يتيح الانحدار الخطي تفسيرًا بسيطًا للعلاقة بين المتغيرات باستخدام معاملات الانحدار.

- الانحدار اللوجستي: قد يكون من الصعب تفسير الانحدار اللوجستي بشكل مباشر، حيث يتم تقديم النتائج عادة على شكل احتماليات الحدوث.

باختصار، الانحدار الخطي واللوجستي يختلفان في الهدف والتوزيع والاستخدامات، حيث يُستخدم الانحدار الخطي لتحليل البيانات المستمرة وتوقع القيم، بينما يُستخدم الانحدار اللوجستي لتحليل البيانات الثنائية وتوقع احتمالات الحدوث للحدث الثنائي.

(٢.٢) المقارنة بين الأداء الإحصائي للانحدار الخطي والانحدار اللوجستي:

١ .صلة العلاقة:

- الانحدار الخطي يقيم العلاقة بين متغير تابع مستمر ومتغيرات مستقلة، بينما الانحدار اللوجستي يقيم العلاقة بين متغير تابع ثنائي القيمة ومتغيرات مستقلة.

٢ .التوزيع:

- يفترض الانحدار الخطي توزيعًا طبيعيًا للبيانات، بينما يفترض الانحدار اللوجستي توزيع احتمالات الحدوث .

٣ .التوقعات:

- في الانحدار الخطي، يتم توقع القيم المستقبلية للمتغير التابع، بينما في الانحدار اللوجستي، يتم توقع احتمالات الحدوث للحدث الثنائي .

٥ .التفسير:

- قد يكون التفسير بسيطًا أكثر في الانحدار الخطي حيث يمكن فهم المعاملات بسهولة، بينما قد يكون التفسير أكثر تعقيدًا في الانحدار اللوجستي بسبب استخدام الاحتمالات.

باختصار، كل من الانحدار الخطي والانحدار اللوجستي لهما مزايا واستخداماتهما الخاصة، ويعتمد اختيار النموذج المناسب على السياق البحثي والأهداف المحددة للتحليل.

(٣٢.) الاستخدام العملي لكل من الانحدار الخطي واللوجستي:

١ .تحليل العلاقة بين الإعلانات والمبيعات:

- استخدام الانحدار الخطي: يمكن استخدام الانحدار الخطي لتحليل العلاقة بين الإعلانات المنفقة عليها والمبيعات الناتجة، مما يمكن أن يساعد في تقدير تأثير الإعلانات على المبيعات.

- استخدام الانحدار اللوجستي: يمكن استخدام الانحدار اللوجستي لتحليل احتمالات الشراء بناءً على عوامل مثل الإعلانات والعوامل الاقتصادية الأخرى.

٢ .تقدير احتمالات الإصابة بالمرض:

- استخدام الانحدار اللوجستي: يمكن استخدام الانحدار اللوجستي لتحليل عوامل الخطر وتقدير احتمالات الإصابة بمرض معين، مثل تحليل عوامل الخطر لسرطان الثدي.

- استخدام الانحدار الخطي: قد يكون من الصعب استخدام الانحدار الخطي في هذه الحالة لأن النتيجة المطلوبة هي احتمالية الحدوث وليست قيمة مستمرة.

٣ .توقع الاستجابة للعلاج:

- استخدام الانحدار اللوجستي: يمكن استخدام الانحدار اللوجستي لتحليل عوامل التوقعات المتعددة وتقدير احتمالات الاستجابة لعلاج معين، مثل توقع الاستجابة للأدوية.

- استخدام الانحدار الخطي: يمكن استخدام الانحدار الخطي لتحليل تأثير العوامل المستقلة المستمرة على النتائج المستمرة، مثل تقدير تأثير جرعة معينة من الدواء على ضغط الدم.

٤ .تحليل النمو السكاني:

- استخدام الانحدار الخطي: يمكن استخدام الانحدار الخطي لتحليل العلاقة بين عوامل مثل الدخل والتعليم وعدد الأطفال في الأسرة ونمو السكان.

- استخدام الانحدار اللوجستي: يمكن استخدام الانحدار اللوجستي لتحليل احتمالات الانتقال بين فئات السن بناءً على العوامل الاجتماعية والاقتصادية.

الانحدار الخطي والانحدار اللوجستي يتم استخدامهما في سياقات مختلفة تبعًا للهدف المحدد للتحليل ونوع البيانات المتاحة، وكل منهما يوفر رؤى مختلفة حول العلاقة بين المتغيرات.

الفصل الثالث:

كيف يعمل الانحدار الخطي واللوجستي في التعلم الآلي مع تطبيق عملي في Python

الفصل الثالث:

كيف يعمل الانحدار الخطي واللوجستي في التعلم الآلي مع تطبيق عملي في Python

(١.٣) مقدمة:

يعتبر كل من الانحدار الخطي والانحدار اللوجستي من أهم خوارزميات التعلم الآلي التي تعد جزءًا من نماذج التعلم الخاضع للإشراف. ونظرًا لكونهما جزء من نماذج التعلم الآلي الخاضع للإشراف، لذا فهما يستفيدان من البيانات الموسومة labeled data لإجراء التنبؤات.

يستخدم الانحدار الخطي للتنبؤ بالقيم المستمرة بينما يمكن استخدام الانحدار اللوجستي في للتنبؤ بالقيم الفئوية والمستمرة على حد سواء، ولكنه يستخدم على نطاق واسع كخوارزمية تصنيف. تهدف نماذج الانحدار إلى توقع القيمة بناءً على ميزات مستقلة.

الاختلاف الرئيسي الذي يجعل كلاهما مختلفين عن بعضهما البعض هو عندما تكون المتغيرات التابعة ثنائية أو ثلاثية بمعنى أخر فئوية فعندئذ يتم استخدام الانحدار اللوجستي أما عندما تكون المتغيرات التابعة مستمرة، عندئذ يتم استخدام الانحدار الخطي.

(٢.٣) الانحدار الخطي في تعلم الالة

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.datasets import load_boston
from sklearn.model_selection import train_test_split
boston = load_boston()
boston.data.shape, boston.target.shape
>> output   ((506, 13), (506,))
```

هو نفس النموذج المستخدم على نطاق واسع في التحليل التنبئي .Predictive analysis هذا النموذج يوضح لنا بشكل رئيسي عن العلاقة بين الهدف target متغير تابع dependent variable والميزات predictors باستخدام خط مستقيم. الانحدار الخطي قد يكون انحدار خطي بسيط وانحدار خطي متعدد. بحيث أن الانحدار الخطي البسيط يحتوي على متغير مستقل واحد فقط بينما الانحدار المتعدد يمكن أن يكون هناك أكثر من متغير مستقل واحد.

لنفهم طريقة عمل النموذج هذا دعونا نطبق مسألة من مسائل الانحدار. سنستخدم في هذا المثال مجموعة بيانات بوسطن Boston dataset من مكتبة Scikit-Learn ، تحتوي مجموعة البيانات هذه على معلومات حول قيمة المنزل لمنازل مختلفة في مدينة بوسطن. المتغيرات الأخرى الموجودة في مجموعة البيانات هذه هي معدل الجريمة، مناطق الأعمال غير التجارية في المدينة(INDUS) ، بالإضافة إلى متغيرات أخرى.

أولا سنقوم باستيراد جميع المكتبات المهمة والأهم من ذلك، استيراد مجموعة البيانات من مكتبة .sklearn. datasets

ثانيا سنقوم بعرض مجموعة البيانات بيانيا بمساعدة مكتبة pandas، وسنقوم بتسمية ميزات

```
X_train, X_test, y_train, y_test = train_test_split(boston.data,
                                    boston.target, test_size=0.2)
print(X_train.shape, X_test.shape, y_train.shape, y_test.shape)
>> output (404, 13) (102, 13) (404,) (102,)
```

مجموعة البيانات وبعد ذلك، سنقوم بإنشاء إطار بيانات باستخدام مكتبة pandas.

ثالثا في هذه الخطوة، سنقسم مجموعة البيانات إلى مجموعة تدريب ومجموعة اختبار.

رابعا نقوم بملائمة مجموعة البيانات الخاصة بنا مع نموذج التعلم الآلي من sklearn لتنفيذ

```
boston.feature_names
bos = pd.DataFrame(boston.data)
bos.columns = boston.feature_names
print(bos.head())
>> output
   CRIM    ZN   INDUS CHAS NOX    ... RAD TAX    PTRATIO B L     STAT
```

```
print("Train:", sklinreg.score(X_train, y_train))
print("Test:", sklinreg.score(X_test, y_test))
>> output Train: 0.7221378898881832
          Test: 0.8065460951660129
```

الانحدار الخطي.

أخيرا، نقوم باختبار نتائج التدريب والاختبار على مجموعة البيانات الخاصة بنا.

(3.3) الانحدار اللوجستي في تعلم الالة

هي خوارزمية يتم استخدامها على نطاق واسع في مهام التصنيف. يتم استخدامها للتنبؤ بالمتغيرات الفئوية بمساعدة المتغيرات التابعة. أن هناك فئتين ونقطة بيانات جديدة يجب التحقق

من الفئة التي تنتمي إليها. ثم تحسب الخوارزميات قيم الاحتمالات التي تتراوح ما بين ٠ و١. على سبيل المثال، هل ستمطر اليوم أم لا. في الانحدار اللوجستي، يتم تمرير مجموع المدخلات المرجح من خلال sigmoid activation function ويسمى المنحنى الذي يتم الحصول عليه بمنحنى سيقمويد .sigmoid curve الدالة اللوجستية التي هي دالة سيقمويد هي منحنى على شكل حرف "S" يأخذ أي قيم حقيقية ويحولها إلى ٠ أو ١. إذا كان الناتج المعطاة من دالة سيقمويد أكثر من ٠.٥، فسيتم تصنيف الناتج على أنه ١ وإذا كان أقل من ٠.٥، فسيتم تصنيف الناتج على أنه ٠. إذا وصل الرسم البياني إلى سالب مالا نهاية، فإن y المتوقع سيكون ٠ والعكس صحيح.

لفهم الانحدار اللوجستي أكثر نرى المثال التالي على بيانات عشوائية سيتم إنشائها ضمن الخطوات التالية:

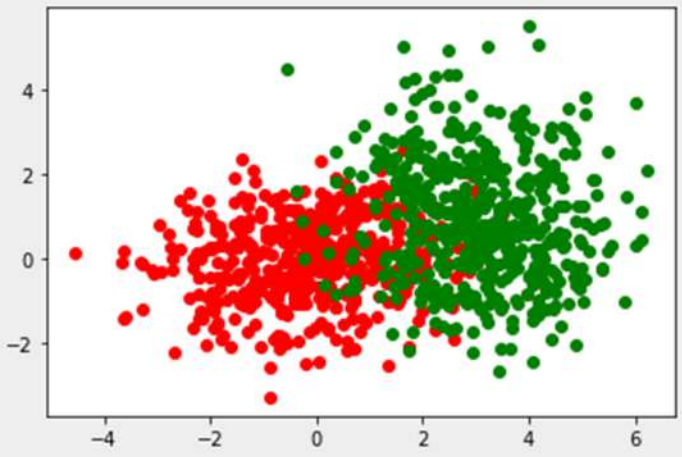
أولا سنقوم باستيراد جميع المكتبات المهمة

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression as SKLR
```

ثانيا نقوم بإنشاء مجموعة بيانات مكونة من ١٠٠٠ صف وعمودين ونقوم برسم مجموعة البيانات هذه بمساعدة مكتبة matplotlib.

```
mean_01 = [0,0]
cov_01 = [[2,0.2], [0.2,1]]
mean_02 = [3,1]
cov_02 = [[1.5,-0.2], [-0.2,2]]
dist_01 = np.random.multivariate_normal(mean_01, cov_01, 500)
dist_02 = np.random.multivariate_normal(mean_02, cov_02, 500)
print(dist_01.shape, dist_02.shape)
>> output (500, 2) (500, 2)
```

```
plt.figure()
plt.scatter(dist_01[:,0], dist_01[:,1], color='red')
plt.scatter(dist_02[:,0], dist_02[:,1], color='green')
plt.show()
>> output
```



ثالثا عند إضافة كلا التوزيعين، تتم إضافة "١" هنا بسبب عمود التسمية

يتم توزيع أول ٥٠٠ نقطة بيانات للتوزيع الأول

أيضا يتم توزيع فيما بعد ٥٠٠ نقطة بيانات للتوزيع الثاني

وعمل عمود منفصل للتسميات.

```
dataset = np.zeros((dist_01.shape[0] + dist_02.shape[0], dist_01.shape[1] + 1))
dataset[:dist_01.shape[0], :-1] = dist_01
dataset[dist_01.shape[0]:, :-1] = dist_02
 # Red = 0, Green = 1
 dataset[dist_02.shape[0]:, -1] = 1
dataset.shape
>> Output (1000, 3)
```

لاحقا يتم تبديل مجموعة البيانات عشوائيًا بحيث يتم خلط التوزيعين بشكل صحيح بحيث يعملان كمجموعة بيانات حقيقية ومن ثم يتم تقسيم مجموعة البيانات لمجموعة تدريب ومجموعة اختبار ثم نرسم الشكل البياني لمجموعة البيانات

```
np.random.shuffle(dataset)
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(dataset[:,:-1], dataset[:,-1], test_size=0.2)
X_train.shape, X_test.shape, y_train.shape, y_test.shape
>> Output ((800, 2), (200, 2), (800,), (200,))
```

أخيرا نقوم بملائمة البيانات على نموذج الانحدار اللوجستي

```
sk_logreg = SKLR()
sk_logreg.fit(X_train, y_train)
>> Output LogisticRegression(C=1.0, class_weight=None, dual=False,
                fit_intercept=True, intercept_scaling=1, l1_ratio=None,
                max_iter=100, multi_class='auto', n_jobs=None, penalty='l2',
                random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                warm_start=False)

sk_logreg.score(X_test, y_test)
>> Output 0.89
```

الفصل الرابع
الاستنتاجات

الفصل الرابع

الاستنتاجات

بناءً على البحث المجرى، يمكن استنتاج أن الانحدار الخطي والانحدار اللوجستي هما نموذجان فعالان في مجال تعلم الآلة وتحليل البيانات. كلاهما يستخدم لحل مشكلات التنبؤ والتصنيف، لكنهما يختلفان في الطرق التي يستخدمانها والنتائج التي يوفرانها.

الانحدار الخطي يعمل على تحديد العلاقة الخطية بين المتغيرات، ويستخدم تقنيات الجبر الخطي لتحديد خط مستقيم يمثل أفضل تقدير للعلاقة. من خلال تدريب النموذج على مجموعة من البيانات، يتم تحديد المعاملات المثلى للخط المستقيم، ويمكن استخدام النموذج للتنبؤ بقيم المتغير المستجيب.

من ناحية أخرى، الانحدار اللوجستي يستخدم للتصنيف ثنائي الفئة، حيث يكون المتغير المستجيب متغيرًا ثنائيًا. يعتمد النموذج في ذلك على الدالة اللوجستية لتحقيق التحويل، ويعطي نتائج بين ٠ و١ التي يمكن تفسيرها كاحتمالات.

الاستنتاج العام هو أن الانحدار الخطي يناسب الحالات التي تتطلب تحليلًا للعلاقة الخطية بين المتغيرات، في حين يناسب الانحدار اللوجستي الحالات التي تتطلب تصنيفًا ثنائيًا دقيقًا. ومع ذلك، يجب اختيار النموذج المناسب وفقًا لطبيعة المشكلة والبيانات المتاحة. قد يتطلب الأمر تجربة عدة نماذج لتحقيق أداء مثلى.

باختصار، الانحدار الخطي والانحدار اللوجستي هما أدوات قوية في تعلم الآلة، ويمكن استخدامهما في مجموعة واسعة من التطبيقات للتنبؤ والتصنيف بدقة.

المصادر

١. كيف يعمل الانحدار الخطي واللوجستي في التعلم الآلي مع تطبيق عملي في - Python علوم ٢٤ (sciences24.com).

٢. تحليل الانحدار | دليل تحليل | موقعQuestioner .

٣. محمد عبد الرحمن إسماعيل: تحليل الانحدار الخطي: الرياض، ٢٠١٦.

٤. د. سعد بن سعيد القحطاني: الإحصاء التطبيقي: الرياض، ٢٠١٥.

٥. تقى الخزعلي: تقدير انموذج انحدار لوجستي لمتغير معتمد متعدد المستويات: رسالة ماجستير،٢٠٢١.

٦. أ. محمد أمين دعيش: نموذج الانحدار اللوجستي: مفهومه، خصائصه، تطبيقاته: جامعة محمد لمين دباغين سطيف – ٢: الجزائر، ٢٠١٧

7. DOUGLAS C. MONTGOMERY: INTRODUCTION TO LINEAR REGRESSION ANALYSIS: Arizona State University, 2021