سلسلة تمارين محلولة في الانحدار اللوجستي

تمرين 1: بناء وتحليل نموذج انحدار لوجستي

لدينا مجموعة بيانات تحتوي على 10 ملاحظات حول طلاب في مدرسة، حيث نريد التنبؤ بما إذا كان الطالب سيحصل على درجة نهائية مرتفعة ($Y = 1$) أم منخفضة ($Y = 0$) بناء على عدد الساعات التي درسها الطالب (X_1) ومعدل الحضو (X_2) . البيانات كالتالي:

الطالب	X_1 (عدد الساعات)	X_2 (معدل الحضور)	Y (النتيجة)
1	5	80	0
2	7	90	1
3	6	70	0
4	8	95	1
5	4	60	0
6	9	85	1
7	3	50	0
8	10	90	1
9	2	40	0
10	7	75	1

المطلوب:

1. بناء نموذج انحدار لوجستي باستخدام البيانات السابقة.
2. إذا علمت أن $\beta_0 \approx -15.7, \quad \beta_1 \approx 1.5, \quad \beta_2 \approx 0.15$
   قم بصياغة النموذج المعدل وحساب احتمال النجاح لكل طالب.
3. قم بتقييم أداء النموذج باستخدام مصفوفة الارتباك وحساب الدقة والدقة التنبؤية والاستدعاء.

الحل

1. صياغة النموذج

نستخدم معادلة الانحدار اللوجستي:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}$$

حيث:

- $P(Y = 1)$: احتمال الحصول على نتيجة عالية.
- β_0: المقطع الثابت.

- β_1: معامل عدد الساعات (X_1).
- β_2: معامل معدل الحضور (X_2).

صياغة النموذج المعدل:

$$P(Y = 1) = \frac{1}{1 + e^{-(-15{,}7+1{,}5X_1+0{,}15X_2)}}$$

2. حساب احتمال $P(Y = 1)$ لكل طالب

نستخدم النموذج المقدر لحساب احتمال كل طالب:

$$P(Y = 1) = \frac{1}{1 + e^{-(-15{,}7+1{,}5X_1+0{,}15X_2)}}$$

حساب الاحتمالات لكل طالب:

حساب احتمال النجاح لكل طالب

نحسب الاحتمال $P(Y = 1)$ لكل طالب باستخدام المعادلة:

الطالب	X_1	X_2	$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2$	$P(Y = 1)$	التوقع ($\hat{Y}$)	(Y الفعلية)	الفئة
1	5	80	$-15.7 + 1.5(5) + 0.15(80) = 3.8$	0.978	1	0	**FP**
2	7	90	$-15.7 + 1.5(7) + 0.15(90) = 8.3$	1.000	1	1	**TP**
3	6	70	$-15.7 + 1.5(6) + 0.15(70) = 3.8$	0.978	1	0	**FP**
4	8	95	$-15.7 + 1.5(8) + 0.15(95) = 10.55$	1.000	1	1	**TP**
5	4	60	$-15.7 + 1.5(4) + 0.15(60) = -0.7$	0.332	0	0	**TN**
6	9	85	$-15.7 + 1.5(9) + 0.15(85) = 10.55$	1.000	1	1	**TP**
7	3	50	$-15.7 + 1.5(3) + 0.15(50) = -3.7$	0.024	0	0	**TN**
8	10	90	$-15.7 + 1.5(10) + 0.15(90) = 12.8$	1.000	1	1	**TP**
9	2	40	$-15.7 + 1.5(2) + 0.15(40) = -6.7$	0.001	0	0	**TN**
10	7	75	$-15.7 + 1.5(7) + 0.15(75) = 6.05$	0.997	1	1	**TP**

ملاحظة: العمود الأخير يُصنف كل تنبؤ إلى واحدة من أربع فئات:

(True Positive) TP: النموذج توقع "نجاح" والقيمة الفعلية "نجاح" (مثل الطلاب 2، 4، 6، 8، 10).

(True Negative) TN: النموذج توقع "فشل" والقيمة الفعلية "فشل" (مثل الطلاب 5، 7، 9).

(False Positive) FP: النموذج توقع "نجاح" بينما القيمة الفعلية "فشل" (مثل الطلاب 1، 3).

(False Negative) FN: النموذج توقع "فشل" بينما القيمة الفعلية "نجاح" (لا يوجد في الجدول).

على سبيل المثال، الطالب 1 لديه احتمال نجاح عالٍ (0.978) رغم أن نتائجه الفعلية كانت فشلًا (Y=0)، مما يؤدي إلى توقع خاطئ (FP). في المقابل، الطالب 2 لديه احتمال نجاح (1.000) ونتيجة فعلية (1)، مما يجعل التنبؤ صحيحًا (TP). الجدول يُظهر أن النموذج حقق 5 تنبؤات صحيحة بـ "نجاح" (TP)، و3 تنبؤات صحيحة بـ "فشل" (TN)، و2 تنبؤات خاطئة بـ "نجاح" (FP)، دون أي تنبؤات خاطئة بـ "فشل" (FN=0)، مما يدل على دقة عالية في تصنيف النجاحات.

4. تقييم أداء النموذج

تحديد الحدود:(Threshold)

نستخدم حدا عتبة قدره 0.5 لتحديد التنبؤ:

- إذا $P(Y = 1) > 0.5$, فإن التنبؤ هو $Y = 1$.
- إذا $P(Y = 1) \leq 0.5$, فإن التنبؤ هو $Y = 0$.

TN = 3, FP = 2, FN = 0, TP = 5

حساب المقاييس:

- **الدقة (Accuracy)**:

$$\frac{TP + TN}{Total} = \frac{5 + 3}{10} = 0.8\ (80\%)$$

- **الدقة التنبؤية (Precision)**:

$$\frac{TP}{TP + FP} = \frac{5}{5 + 2} \approx 0.714\ (71.4\%)$$

- **الاستدعاء (Recall)**:

$$\frac{TP}{TP + FN} = \frac{5}{5 + 0} = 1\ (100\%)$$

تمرين 2: تحليل بيانات العملاء باستخدام الانحدار اللوجستي

لدينا مجموعة بيانات تحتوي على معلومات حول عملاء شركة اتصالات، حيث نريد التنبؤ بما إذا كان العميل سيترك الشركة ($Y = 1$) (churn) أم سيبقى ($Y = 0$) بناء على عدد الدقائق التي يتحدثها العميل شهريا (X_1)ومقدار الفاتورة الشهرية .(X_2) البيانات كالتالي:

العميل	X_1 (عدد الدقائق)	X_2 (الفاتورة الشهرية بالدولار)	Y (الحالة)
1	200	50	0
2	300	70	1
3	250	60	0
4	350	80	1
5	150	40	0
6	400	90	1
7	100	30	0
8	450	100	1
9	50	20	0
10	300	65	1

المطلوب:

1. بناء نموذج انحدار لوجستي باستخدام هذه البيانات.
2. كتابة النموذج وحساب احتمال مغادرة العميل ($P(Y = 1)$) لكل عميل، إذا علمت أن:

$$\beta_0 = -5.0, \quad \beta_1 = 0.01, \quad \beta_2 = 0.05$$

3. تقييم أداء النموذج عن طريق حساب الدقة والدقة التنبؤية والاستدعاء.

الحل

1. بناء النموذج اللوجستي

صياغة النموذج:

النموذج اللوجستي يأخذ الشكل التالي:

$$P(Y = 1) = \frac{1}{1 + e^{-(-5+0.01X_1+0.05X_2)}}$$

حيث:

- $P(Y = 1 \mid X_1, X_2)$ هو احتمال أن يترك العميل الشركة.

2. حساب احتمال $P(Y = 1)$ لكل عميل

نستخدم النموذج المقدر لحساب احتمال كل عميل:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0+\beta_1X_1+\beta_2X_2)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5+0.01X_1+0.05X_2)}}$$

حساب الاحتمالات لكل عميل:

العميل الأول: $(X_1 = 200, X_2 = 50)$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+0.01\cdot200+0.05\cdot50)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+2.0+2.5)}} = \frac{1}{1 + e^{-0.5}}$$

$$P(Y = 1) \approx 0.38$$

العميل الثاني: $(X_1 = 300, X_2 = 70)$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+0.01\cdot300+0.05\cdot70)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+3.0+3.5)}} = \frac{1}{1 + e^{1.5}}$$

$$P(Y = 1) \approx 0.82$$

العميل الثالث: $(X_1 = 250, X_2 = 60)$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+0.01\cdot250+0.05\cdot60)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+2.5+3.0)}} = \frac{1}{1 + e^{0.5}}$$

$$P(Y = 1) \approx 0.62$$

العميل الرابع: $(X_1 = 350, X_2 = 80)$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+0.01\cdot350+0.05\cdot80)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+3.5+4.0)}} = \frac{1}{1 + e^{2.5}}$$

$$P(Y = 1) \approx 0.92$$

العميل الخامس: $(X_1 = 150, X_2 = 40)$

$$P(Y = 1) = \frac{1}{1 + e^{-(-5.0+0.01\cdot150+0.05\cdot40)}}$$

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+1.5+2.0)}} = \frac{1}{1+e^{-1.5}}$$
$$P(Y=1) \approx 0.18$$

العميل السادس:$(X_1 = 400, X_2 = 90)$

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 400+0.05\cdot 90)}}$$
$$P(Y=1) = \frac{1}{1+e^{-(-5.0+4.0+4.5)}} = \frac{1}{1+e^{3.5}}$$
$$P(Y=1) \approx 0.97$$

العميل السابع:$(X_1 = 100, X_2 = 30)$

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 100+0.05\cdot 30)}}$$
$$P(Y=1) = \frac{1}{1+e^{-(-5.0+1.0+1.5)}} = \frac{1}{1+e^{-2.5}}$$
$$P(Y=1) \approx 0.08$$

العميل الثامن:$(X_1 = 450, X_2 = 100)$

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 450+0.05\cdot 100)}}$$
$$P(Y=1) = \frac{1}{1+e^{-(-5.0+4.5+5.0)}} = \frac{1}{1+e^{4.5}}$$
$$P(Y=1) \approx 0.99$$

العميل التاسع:$(X_1 = 50, X_2 = 20)$

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 50+0.05\cdot 20)}}$$
$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.5+1.0)}} = \frac{1}{1+e^{-3.5}}$$
$$P(Y=1) \approx 0.03$$

العميل العاشر:$(X_1 = 300, X_2 = 65)$

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 300+0.05\cdot 65)}}$$
$$P(Y=1) = \frac{1}{1+e^{-(-5.0+3.0+3.25)}} = \frac{1}{1+e^{1.25}}$$
$$P(Y=1) \approx 0.78$$

جدول الاحتمالات:

العميل	X_1 (الدقائق)	X_2 (الفاتورة)	$P(Y=1)$	$\hat{Y}$ (التوقع)	Y (الفعلية)	الفئة
1	200	50	0.38	0	0	TN
2	300	70	0.82	1	1	TP
3	250	60	0.62	1	0	FP
4	350	80	0.92	1	1	TP
5	150	40	0.18	0	0	TN
6	400	90	0.97	1	1	TP
7	100	30	0.08	0	0	TN
8	450	100	0.99	1	1	TP
9	50	20	0.03	0	0	TN
10	300	65	0.78	1	1	TP

3. تقييم أداء النموذج

تحديد الحدود:(Threshold)

نستخدم حدا عتبة قدره 0.5 لتحديد التنبؤ:

- إذا $P(Y=1) > 0.5$, فإن التنبؤ هو $Y=1$.
- إذا $P(Y=1) \leq 0.5$, فإن التنبؤ هو $Y=0$.

مصفوفة الارتباك:

- True Positives = 5
- False Positives = 1
- True Negatives = 4
- False Negatives = 0

حساب المقاييس:

الدقة (Accuracy):

$$\text{Accuracy} = \frac{TP + TN}{\text{Total}} = \frac{5+4}{10} = 0.9\ (90\%)$$

التنبؤية الدقة (Precision):

$$\text{Precision} = \frac{TP}{TP + FP} = \frac{5}{5+1} \approx 0.833\ (83.3\%)$$

الاستدعاء (Recall):

$$\text{Recall} = \frac{TP}{TP + FN} = \frac{5}{5+0} = 1.0\ (100\%)$$

تمرين 3: تحليل بيانات المرضى باستخدام الانحدار اللوجستي

لدينا مجموعة بيانات تحتوي على معلومات حول مرضى في مستشفى، حيث نريد التنبؤ بما إذا كان المريض سيحتاج إلى دخول العناية المركزة ($Y = 1$) أم لا ($Y = 0$) بناء على العمر (X_1) ومستوى السكر في الدم (X_2).البيانات كالتالي

:

المريض	X_1 (العمر)	X_2 (مستوى السكر)	Y (الحالة)
1	45	120	0
2	60	150	1
3	50	130	0
4	70	180	1
5	35	110	0
6	65	160	1
7	40	100	0
8	55	140	1

المطلوب:

1. بناء نموذج انحدار لوجستي باستخدام هذه البيانات.
2. إذا علمت أن

$$\beta_0 = -10.0, \quad \beta_1 = 0.1, \quad \beta_2 = 0.05$$

أكتب النموذج المعدل

3. حساب احتمال الحاجة إلى دخول العناية المركزة ($P(Y = 1)$) لكل مريض.
4. تقييم أداء النموذج عن طريق حساب الدقة والدقة التنبؤية والاستدعاء.

الحل

الخطوة 1: صياغة النموذج

نستخدم معادلة الانحدار اللوجستي:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}$$

حيث:

- $P(Y = 1)$: احتمال الحاجة إلى دخول العناية المركزة.
- β_0: المقطع الثابت.
- β_1: معامل العمر.(X_1)
- β_2: معامل مستوى السكر.(X_2)

الخطوة 2: حساب الاحتمالات

نستخدم المعادلة:

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}$$

لحساب احتمال الحاجة إلى دخول العناية المركزة لكل مريض .

للمريض: $(X_1 = 45, X_2 = 120)$ **1**

$$P(Y=1) = \frac{1}{1 + e^{-(-10.0 + 0.1(45) + 0.05(120))}} = \frac{1}{1 + e^{-(-10.0 + 4.5 + 6.0)}} = \frac{1}{1 + e^{0.5}} \approx 0.3775$$

للمريض: $(X_1 = 60, X_2 = 150)$ **2**

$$P(Y=1) = \frac{1}{1 + e^{-(-10.0 + 0.1(60) + 0.05(150))}} = \frac{1}{1 + e^{-(-10.0 + 6.0 + 7.5)}} = \frac{1}{1 + e^{3.5}} \approx 0.9707$$

للمرضى الآخرين:

نتبع نفس الخطوات لحساب الاحتمالات .النتائج ملخصة في الجدول أدناه:

المريض	$P(Y=1)$
1	0.6225
2	0.9705
3	0.8176
4	0.9975
5	0.2689
6	0.9890
7	0.2689
8	0.9240

الخطوة 3: تصنيف المرضى

نستخدم عتبة $(P > 0.5)$ لتصنيف المرضى:

- إذا $P(Y=1) > 0.5: Y = 1$ احتياج إلى العناية المركزة.(
- إذا $P(Y=1) \leq 0.5: Y = 0$ لا حاجة إلى العناية المركزة.(

النتائج التصنيفية هي:

المريض	Y_{actual}	$Y_{predicted}$	الفئة
1	0	1	FP
2	1	1	TP
3	0	1	FP
4	1	1	TP
5	0	0	TN
6	1	1	TP
7	0	0	TN
8	1	1	TP

الخطوة :5 تقييم النموذج

حساب المقاييس:

1. **الدقة:**

$$\text{الدقة} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{4 + 2}{4 + 2 + 2 + 0} = \frac{7}{8} = 0.75$$

2. **الدقة التنبؤية:**

$$\text{الدقة التنبؤية} = \frac{TP}{TP + FP} = \frac{4}{4 + 2} = \frac{4}{6} = 0.666$$

3. **الاستدعاء:**

$$\text{الاستدعاء} = \frac{TP}{TP + FN} = \frac{4}{4 + 0} = 1.0$$