سلسلة تمارين حول التحليل 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 $(Y=0)$ بناءً 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 بحساب 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).

3.حساب احتمال $P(Y = 1)$ لكل طالب

نستخدم النموذج المقدر لحساب احتمال كل طالب:

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

حساب الاحتمالات لكل طالب :

الطالب الأول:($X_1 = 5$, $X_2 = 80$)

$$P(Y = 1) = \frac{1}{1 + e^{-(-15.7 + 1.5 \cdot 5 + 0.15 \cdot 80)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-15.7 + 7.5 + 12)}} = \frac{1}{1 + e^{-0.2}}$$

$$P(Y = 1) \approx 0.45$$

الطالب الثاني:($X_1 = 7$, $X_2 = 90$)

$$P(Y = 1) = \frac{1}{1 + e^{-(-15.7 + 1.5 \cdot 7 + 0.15 \cdot 90)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-15.7 + 10.5 + 13.5)}} = \frac{1}{1 + e^{-1.7}}$$

$$P(Y = 1) \approx 0.85$$

الطالب الثالث:($X_1 = 6$, $X_2 = 70$)

$$P(Y = 1) = \frac{1}{1 + e^{-(-15.7 + 1.5 \cdot 6 + 0.15 \cdot 70)}}$$

$$P(Y = 1) = \frac{1}{1 + e^{-(-15.7 + 9 + 10.5)}} = \frac{1}{1 + e^{-2.8}}$$

$$P(Y = 1) \approx 0.37$$

الطالب الرابع:($X_1 = 8$, $X_2 = 95$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 8+0.15\cdot 95)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+12+14.25)}}=\frac{1}{1+e^{-0.55}}$$

$$P(Y=1)\approx 0.95$$

الطالب الخامس:($X_1=4, X_2=60$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 4+0.15\cdot 60)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+6+9)}}=\frac{1}{1+e^{-0.7}}$$

$$P(Y=1)\approx 0.26$$

الطالب السادس:($X_1=9, X_2=85$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 9+0.15\cdot 85)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+13.5+12.75)}}=\frac{1}{1+e^{-0.55}}$$

$$P(Y=1)\approx 0.81$$

الطالب السابع:($X_1=3, X_2=50$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 3+0.15\cdot 50)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+4.5+7.5)}}=\frac{1}{1+e^{-3.7}}$$

$$P(Y=1)\approx 0.18$$

الطالب الثامن:($X_1=10, X_2=90$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 10+0.15\cdot 90)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+15+13.5)}}=\frac{1}{1+e^{-2.8}}$$

$$P(Y=1)\approx 0.98$$

الطالب التاسع:($X_1=2, X_2=40$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 2+0.15\cdot 40)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+3+6)}}=\frac{1}{1+e^{-6.7}}$$

$$P(Y=1)\approx 0.12$$

الطالب العاشر:($X_1=7, X_2=75$)

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+1.5\cdot 7+0.15\cdot 75)}}$$

$$P(Y=1)=\frac{1}{1+e^{-(-15.7+10.5+11.25)}}=\frac{1}{1+e^{-0.55}}$$

$$P(Y=1)\approx 0.78$$

جدول الاحتمالات :

الطالب	$P(Y=1)$
1	0.45
2	0.85
3	0.37
4	0.95
5	0.26
6	0.81
7	0.18
8	0.98
9	0.12
10	0.78

4. تقييم أداء النموذج

أ- تحديد الحدود :(Threshold)

نستخدم حدًا عتبة قدره 0.5 لتحديد التنبؤ:

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

ب- بناء مصفوفة الارتباك :

الواقع/التنبؤ	$Y=0$	$Y=1$
$Y=0$	4	1
$Y=1$	1	4

ج- حساب المقاييس :

الدقة:(Accuracy)

$$\text{Accuracy} = \frac{\text{عدد التنبؤات الصحيحة}}{\text{إجمالي الملاحظات}} = \frac{4+4}{10} = 0.8$$

الدقة التنبؤية:(Precision)

$$\text{Precision} = \frac{\text{عدد الإيجابيات الحقيقية}}{\text{إجمالي التنبؤات الإيجابية}} = \frac{4}{4+1} = 0.8$$

الاستدعاء:(Recall)

$$\text{Recall} = \frac{\text{عدد الإيجابيات الحقيقية}}{\text{إجمالي الحالات الإيجابية الواقعية}} = \frac{4}{4+1} = 0.8$$

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 $(Y = 0)$ بناءً على عدد الدقائق التي يتحدثها العميل شهريًا (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:
3. $\beta_0 = -5.0, \quad \beta_1 = 0.01, \quad \beta_2 = 0.05$
4. تقييم أداء النموذج باستخدام مصفوفة الارتباك وحساب الدقة والدقة التنبؤية والاستدعاء.

الحل التفصيلي

1. بناء النموذج اللوجستي

صياغة النموذج:

النموذج اللوجستي يأخذ الشكل التالي:

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

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:

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

حساب الاحتمالات لكل عميل:

العميل الأول:($X_1 = 200$, $X_2 = 50$)

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 200+0.05\cdot 50)}}$$

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

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

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

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 300+0.05\cdot 70)}}$$

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

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

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

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 250+0.05\cdot 60)}}$$

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

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

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

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 350+0.05\cdot 80)}}$$

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

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

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

$$P(Y=1) = \frac{1}{1+e^{-(-5.0+0.01\cdot 150+0.05\cdot 40)}}$$

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

$$P(Y=1) \approx 0.18$$

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	$P(Y=1)$
1	0.38
2	0.82
3	0.62
4	0.92
5	0.18
6	0.97
7	0.08
8	0.99
9	0.03
10	0.78

3. تقييم أداء النموذج

أ- تحديد الحدود:(Threshold)

نستخدم حدًا 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 $Y=0$.

ب- بناء مصفوفة الارتباك:

المتنبؤ به/الحقيقي	$Y=0$	$Y=1$
$Y=0$	4	1
$Y=1$	1	4

ج- حساب المقاييس:

1. **الدقة (Accuracy)**

$$\text{Accuracy}=\frac{\text{عدد التنبؤات الصحيحة}}{\text{إجمالي الملاحظات}}=\frac{4+4}{10}=0.8$$

2. **الدقة التنبؤية (Precision)**

$$\text{Precision}=\frac{\text{عدد الإيجابيات الحقيقية}}{\text{إجمالي التنبؤات الإيجابية}}=\frac{4}{4+1}=0.8$$

3. الاستدعاء **(Recall)**

$$\text{Recall} = \frac{\text{عدد الإيجابيات الحقيقية}}{\text{إجمالي الحالات الإيجابية الحقيقية}} = \frac{4}{4+1} = 0.8$$

- **معادلة النموذج:**

$$P(Y = 1 \mid X_1, X_2) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2)}}, \quad \text{حيث } \beta_0 = -5.0, \beta_1 = 0.01, \beta_2 = 0.05$$

- **مصفوفة الارتباك:**
  - True Positives = 4
  - False Positives = 1
  - True Negatives = 4
  - False Negatives = 1

- **مقاييس الأداء:**
  - Accuracy = $\boxed{0.8}$
  - Precision = $\boxed{0.8}$
  - Recall = $\boxed{0.8}$

تمرين3: 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 $(Y = 0)$ بناءً 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 باستخدام مصفوفة الارتباك وحساب الدقة والدقة التنبؤية والاستدعاء.

الحل التفصيلي

الخطوة 1 صياغة النموذج

نستخدم معادلة الانحدار اللوجستي:

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

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 تقدير المعاملات باستخدام MLE

لإيجاد المعاملات $(\beta_0,\beta_1,\beta_2)$ ، نقوم بتعظيم دالة الاحتمال الأقصى:

$$\ln L(\beta)=\sum_{i=1}^{N}[Y_i\ln P_i+(1-Y_i)\ln(1-P_i)]$$

الخطوة 1.2 اختيار قيم أولية للمعاملات

نبدأ بقيم أولية:

$$\beta_0=0,\quad \beta_1=0,\quad \beta_2=0$$

الخطوة 2.2: تحديث المعاملات باستخدام خوارزمية التحسين

نستخدم خوارزمية مثل **نيوتن-رافسون** أو **الانحدار التدريجي** لتحديث المعاملات .بعد عدة تكرارات) غير موضحة هنا للتبسيط(، نحصل على القيم التالية:

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

الخطوة 3: حساب الاحتمالات

نستخدم المعادلة:

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

لحساب احتمال الحاجة إلى دخول العناية المركزة لكل مريض .على سبيل المثال:

للمريض 1: $(X_1=45, X_2=120)$

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

للمريض 2: $(X_1 = 60, X_2 = 150)$

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

للمرضى الآخرين:

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

المريض	$P(Y=1)$
1	0.3775
2	0.9707
3	0.6225
4	0.9933
5	0.1824
6	0.9997
7	0.0474
8	0.8176

الخطوة :4تصنيف 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}	$Y_{\text{predicted}}$
1	0	0
2	1	1
3	0	1
4	1	1
5	0	0
6	1	1
7	0	0
8	1	1

الخطوة :5تقييم النموذج

مصفوفة الارتباك:

	$Y_{\text{predicted}} = 0$	$Y_{\text{predicted}} = 1$
$Y_{\text{actual}} = 0$	4	1
$Y_{\text{actual}} = 1$	0	3

حساب المقاييس:

1. **الدقة:**

$$\text{الدقة} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{3 + 4}{3 + 4 + 1 + 0} = \frac{7}{8} = 0.875$$

2. **الدقة التنبؤية:**

$$\text{الدقة التنبؤية} = \frac{TP}{TP + FP} = \frac{3}{3 + 1} = \frac{3}{4} = 0.75$$

3. **الاستدعاء:**

$$\text{الاستدعاء} = \frac{TP}{TP + FN} = \frac{3}{3 + 0} = 1.0$$