

Assignment-5

(1) صح أو خطأ مع شرح الإجابة.

A. الافتراض الأساسي للنموذج (Naïve Bayes) هو أن كل ميزة مستقلة. صحيح. يفترض (Naïve Bayes) أن الميزات المستخدمة لوصف المراقبة هي مستقلة بناءً على الفئة. على الرغم من هذا الافتراض يؤدي نموذج (Naïve Bayes) غالبًا إلى أداء جيد في الممارسة.

B. نظرية (Naïve Bayes) توفر وسيلة لحساب الاحتمال الشرطي $P(A|B)$ من $P(B)$ و $P(A)$ و $P(B|A)$. صحيح. نظرية (Naïve Bayes) هي نظرية أساسية في الاحتمالات تصف احتمال حدوث حدث، بناءً على المعرفة السابقة بالشروط التي قد تكون متعلقة بالحدث. الصيغة هي $P(A|B) = [(P(B|A) * P(A)) \setminus P(B)]$ حيث:

$P(A|B)$ هو الاحتمال الشرطي، $P(B|A)$ هو الاحتمالية،
 $P(A)$ هو الاحتمال السابق (احتمال وقوع الحدث قبل رؤية الدليل)، و $P(B)$ هو الدليل.

C. لديك $P(B|A)$ هل تستطيع حساب $P(A|B)$. خطأ. لحساب $P(A|B)$ ، تحتاج إلى معرفة $P(A)$ (الاحتمال السابق لـ A) و $P(B)$ (الدليل). وجود $P(B|A)$ وحده لا يكفي.

D. إذا كان لديك نموذج ذو دقة عالية (High Accuracy) على مجموعة البيانات التدريبية، يمكنك التأكد بنسبة 100% أن نموذجك جيد للغاية.

خطأ. الدقة العالية على مجموعة البيانات التدريبية لا تضمن أن يؤدي النموذج بشكل جيد على بيانات جديدة غير مرئية (Generalization). يُعتبر Overfitting هو مشكلة شائعة حيث يتعلم النموذج البيانات التدريبية بشكل جيد لكنه يفشل في التعميم إلى بيانات جديدة. يمكنك استعمال Precision and recall ومن ثم F1-score.

(2) كيف يعمل (Naïve Bayes)؟

Naive Bayes هي خوارزمية احتمالية للتعليم الآلي تعتمد على نظرية Bayes. تعتبر الخوارزمية "ساذجة" لأنها تفترض أن جميع الميزات المستخدمة لوصف الملاحظة مستقلة عن بعضها البعض. بالرغم من هذا الافتراض البسيط، إلا أنه غالبًا ما يكون أداء Naive Bayes جيدًا عند الممارسة، خاصة في سيناريوهات تصنيف النصوص وتصفية البريد العشوائي.

فيما يلي نظرة عامة على كيفية عمل Naive Bayes:

1- مبرهنة Bayes:

تعتمد الخوارزمية على نظرية Bayes، التي تصف احتمالية وقوع حدث ما، بناءً على المعرفة المسبقة بالظروف التي قد تكون مرتبطة بالحدث.

2- افتراض الاستقلال (Independence):

يفترض Naive Bayes أن جميع الميزات المستخدمة مستقلة عن بعضها البعض، بالنظر إلى متغير الفئة. وهذا افتراض قوي وغير واقعي في كثير من الأحيان، لكنه يبسط الحسابات.

3- التدريب:

خلال مرحلة التدريب، تحسب الخوارزمية احتمالات (probabilities) كل ميزة في كل فئة. يتم حساب الاحتمالات السابقة لكل فئة بناءً على بيانات التدريب.

4- تنبؤ:

عند إجراء تنبؤات لملاحظة جديدة، تستخدم الخوارزمية نظرية Bayes لحساب احتمالية كل فئة بالنظر إلى الميزات المرصودة. يتم بعد ذلك تعيين الفئة ذات الاحتمالية الأعلى كفتة متوقعة للملاحظة الجديدة.

5- التنعيم (Smoothing):

في خطوة التنعيم في خوارزمية Naive Bayes، يتم تطبيق تقنيات التنعيم للتعامل مع الحالات التي يكون فيها احتمال وجود قيمة معينة للميزة صفر في فئة معينة. يتم ذلك لتجنب حدوث احتمالات صفر ولتحسين متانة النموذج. ومن بين هذه التقنيات تقنية التنعيم باستخدام لابلاس (Laplace smoothing). هذا يعني إضافة قيمة صغيرة إلى عدد الحالات الممكنة لكل فئة ولكل ميزة أثناء حساب الاحتمالات.

6- التطبيقات:

يُستخدم Naive Bayes بشكل شائع في مهام تصنيف النص، وتصفية البريد العشوائي، وتحليل المشاعر، وغيرها من المواقع التي يكون فيها افتراض الاستقلال جيدًا إلى حد معقول.

(3) إذا كانت لدينا مجموعة البيانات التالية، فقم بتطبيق Naive Bayes للتنبؤ بالميزات الجديدة

$X = \{\text{color=orange, Calories=2, Tall=long, Tasty=Bad}\}$

Color	Calories	Tall	Tasty	Fruit
Yellow	3	Short	Good	Y
Orange	2	Long	Bad	Y
Orange	3	Short	Good	Y
Yellow	3	Short	Good	Y
Orange	2	Short	Bad	N
Yellow	2	Long	Bad	N
Yellow	2	Long	Bad	N
Yellow	2	Short	Good	N

الحل:

$$Y = 4/8 = 0.5$$

$$N = 4/8 = 0.5$$

Color	Y	N
Yellow	2/4	3/4
Orange	2/4	1/4

Calories	Y	N
2	1/4	4/4
3	3/4	0/4

Tall	Y	N
Short	3/4	2/4
Long	1/4	2/4

Tasty	Y	N
Bad	1/4	3/4
Good	3/4	1/4

لحساب $X = \{\text{color=orange, Calories=2, Tall=long, Tasty=Bad}\}$

$$P(\text{Fruit} | X) = [P(Y) * P(\text{color=orange}/Y) * P(\text{Calories=2}/Y) * P(\text{Tall=long}/Y) * P(\text{Tasty=Bad}/Y)]$$

$$P(\text{Fruit} | X) = [0.5 * \frac{2}{4} * \frac{1}{4} * \frac{1}{4} * \frac{1}{4}]$$

$$P(\text{Fruit} | X) = 0.00390625$$

$$P(\text{NotFruit} | X) = [P(N) * P(\text{color=orange}/N) * P(\text{Calories=2}/N) * P(\text{Tall=long}/N) * P(\text{Tasty=Bad}/N)]$$

$$P(\text{NotFruit} | X) = [0.5 * \frac{1}{4} * \frac{4}{4} * \frac{2}{4} * \frac{3}{4}]$$

$$P(\text{NotFruit} | X) = 0.046875$$

0.046875 > 0.00390625
 NotFruit > Fruit

إذا الإحتمال الأكبر هو NotFruit

(4) ما هي الفكرة الأساسية وراء SVM؟ ما هو SVM؟

الفكرة الأساسية وراء (Support Vector Machines - SVM) هي البحث عن حدود فاصلة (margin) بين فئتين مختلفتين في البيانات.

الهامش هو المسافة بين حدود القرار وأقرب نقطة من كل فئة في (Dataset). SVM يحاول تحقيق أكبر هامش (Maximum margin) ممكن.

SVM: هو خوارزمية في مجال تعلم الآلي تحت الاشراف تُستخدم لمهام التصنيف والتحليل التنبؤي. تم تطوير SVM للبحث عن حدود فاصلة بين مجموعات من البيانات، تستعمل أكثر في المهام ذات الأبعاد العالية. الخصائص الرئيسية SVM:

1. يستخدم للتصنيف حيث يقوم بفصل البيانات إلى فئتين أو أكثر بناءً على الخصائص المستخدمة.
2. العثور على حدود قرار (decision boundary) بين مجموعات البيانات يكون لديها أكبر هامش (margin) ممكن.
3. الهامش هو المسافة بين حدود القرار وأقرب نقطة من كل فئة. يتم تحسين SVM لتحقيق أكبر هامش ممكن.
4. النواقل هي النقاط التي تقع على حدود القرار أو قريبة منها، وتلعب دورًا حاسمًا في تحديد حدود القرار.
5. التعامل مع البيانات التي لا يمكن تصنيفها بشكل خطي، بفضل استخدام وسائط (kernels) لتحويل البيانات إلى فضاء ذو أبعاد أعلى.
6. تظهر فعالية في التعامل مع مجموعات البيانات التي تحتوي على العديد من (Features) أو الأبعاد.

(5) اشرح المعلمات التي تحتاج إلى تغييرها أو تكيفها في SVM؟

بعض parameters التي تحتاج إلى تغييرها في خوارزمية SVM:

1. معلمة C :

- تُمثل معلمة C قيمة تكلفة التصحيح. إذا كانت قيمة C عالية، فإن النموذج سيحاول تصحيح كل الأخطاء في التصنيف، مما قد يؤدي إلى نموذج معقد. ومن الجيد استخدام قيمة C عالية عندما تكون البيانات معقدة.

2. نوع الكيرنل (Kernel Type):

- يُستخدم الكيرنل لتحويل البيانات إلى فضاء ذو أبعاد أعلى. يمكنك اختيار نوع الكيرنل الذي يناسب نوع البيانات الخاص بك، مثل الكيرنل **linear**، وكيرنل **RBF**، وكيرنل **Polynomial**.

3. معلمة Γ :

- تحكم معلمة Γ في شكل الكيرنل. قيم عالية لـ Γ تعني أن النقاط البعيدة يتم منحها أوزان أقل، مما يؤثر أكثر على حدود القرار. قيم منخفضة لـ Γ تعني أن النقاط البعيدة يتم منحها أوزان أكبر.

ملاحظة: تعديل هذه المعلمات يعتمد على البيانات وطبيعة المشكلة. يُفضل استخدام تقنيات التحقق (Cross-Validation) لاختيار القيم المناسبة لهذه المعلمات أو عن طريق (Tuning the hyper-parameters of an estimator).

(6) افترض أن نموذج LogisticRegression الخاص بنا يتدرب على مجموعتين من بيانات الميزات $(x1, x2)$ ويخرج بالمعادلة التالية : $f(x) = z = a(x1) + b(x2) + c$

حيث:

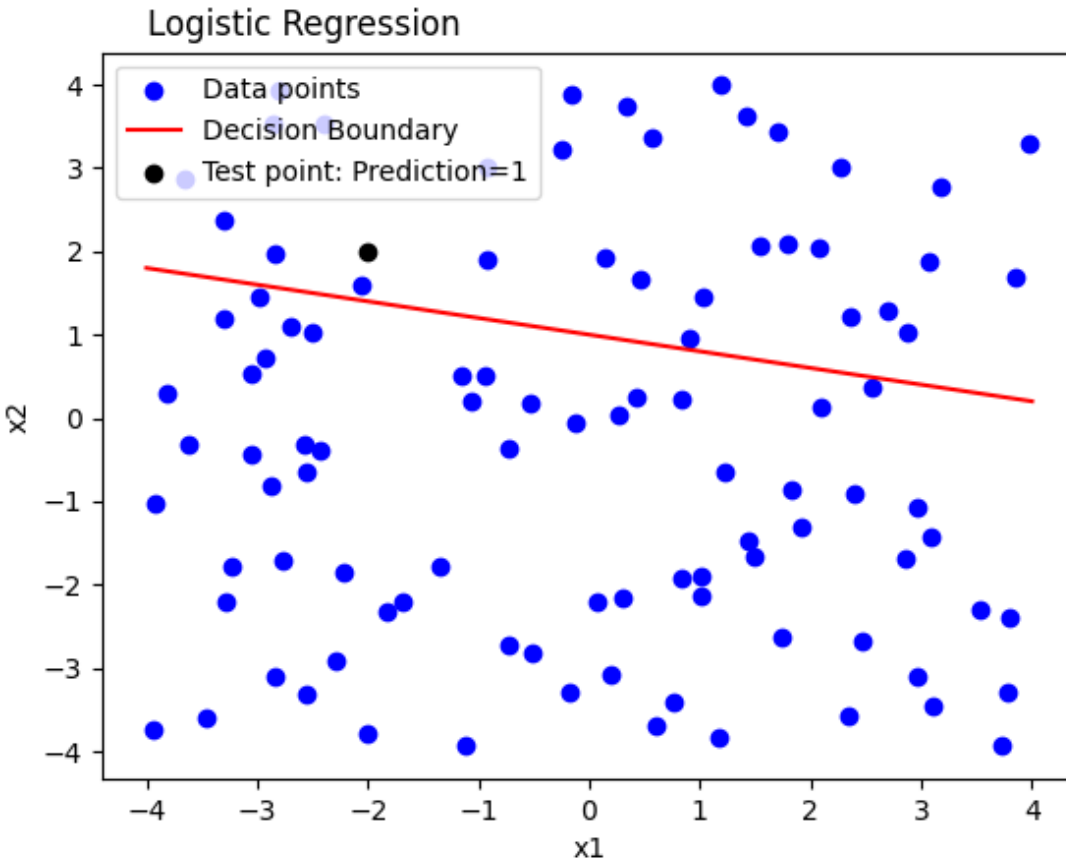
$$a = 3, b = 15, c = -15$$

يتنبأ النموذج بـ $y=1$ إذا كانت $z \geq 0$ ويتنبأ بخلاف ذلك بـ 0.

أ. ارسم بعض البيانات وارسم المنطقة التي يتنبأ فيها النموذج بالرقمين 1 و 0.
ب. ما هو التنبؤ بنموذج عند:

$$x1 = -2$$

$$x2 = 2$$



الداتا ست التي فوق الخط الفاصل تمثل بتوقع 1 (True)
والتي تحت الخط تمثل 0 (False)

التنبؤ عندما $(x1 = -2, x2 = 2)$: هي النقطة السوداء فوق الخط الأحمر وتمثل توقع 1 (True).

(7) افترض أن لدينا مجموعة بيانات وقمنا برسم/تصور مجموعة البيانات كما هو موضح أدناه.
ما هي الخوارزمية التي تقترحها لاستخدام مجموعة البيانات (LR, LogisticRegression, SVM, Naive Bayes)
واشرح إجابتك؟ اشرح أيضاً أي الخوارزميات ستعمل بشكل أفضل على مجموعة البيانات

الحل:

يمكن استخدام Logistic Regression مع تحويلات غير خطية لتعزيز قدرة النموذج على فصل الفئات في حالة وجود حدود غير خطية.

ولكن حسب رأيي الشخصي يعد استعمال Support Vector Machine (SVM) فعال لحالات فصل الفئات حتى في حالة وجود حدود غير خطية. بالإضافة الى استخدام kernel غير خطي (مثل RBF) لتحسين قدرة النموذج على التعامل مع البيانات الغير خطية.