

Classification(KNN)

خولة غيلان+خولة عامر+عائشة الحدادي

علوم حاسوب/ المستوى الرابع

التصنيف (Classification):

هو عملية تستخدم لتصنيف العناصر إلى مجموعات محددة أو فئات بناءً على سماتها أو خصائصها المعطاة. خلال عملية التصنيف، يتم تدريب نموذج التصنيف على مجموعة من البيانات التدريبية التي تتضمن سمات المدخلات والتصنيفات المعروفة. يتم استخدام هذا النموذج المدرب لتصنيف عناصر جديدة غير مصنفة بناءً على سماتها.

KNN:

هو اختصار لـ "k-Nearest Neighbors" وهو خوارزمية تستخدم في تنقيب البيانات و تعلم الآلة (Machine Learning) في مجال التصنيف والتنبؤ. تعتمد الخوارزمية على مفهوم الجيرة وتقوم بتصنيف النقاط في فضاء البيانات بناءً على أقرب جيرة من نقاط التدريب. It is a non-parametric and instance-based learning algorithm.

آلية عمل الخوارزمية:

تعمل خوارزمية KNN على النحو التالي:

1. يتم تدريب النموذج على مجموعة من البيانات المصنفة مسبقًا، تتضمن معلومات حول الأصناف أو التصنيفات.
2. عندما يتم تقديم نقطة جديدة للنموذج للتصنيف، يتم حساب المسافة بين هذه النقطة وجميع النقاط في مجموعة التدريب.
3. يتم تحديد قيمة k ، وهي عدد الجيرة القريبة الذي سيتم اعتباره في التصنيف. من المهم اختيار قيمة مناسبة لـ k حيث يمكن أن يؤثر على دقة وأداء الخوارزمية.
4. يتم تحديد التصنيف الأكثر شيوعًا بين الأصناف المتواجدة في k الجيرة القريبة كتصنيف للنقطة المراد تصنيفها.

مثال:

لنفترض أن لدينا مجموعة من الصور. تم تصنيف كل صورة مسبقًا. عندما يتم تقديم أقرب جيرة من الصور التدريب واخيار التصنيف الأكثر شيوعًا بين تلك الجيرة. على سبيل المثال، إذا كانت $k=3$ وتم تحديد الجيرة الثلاثة الأقرب للصورة الجديدة

Sepal Length	Sepal Width	Species	Distance
5.3	3.7	Setosa	0.608
5.1	3.8	Setosa	0.707
7.2	3.0	Virginica	2.002
5.4	3.4	Setosa	0.36
5.1	3.3	Setosa	0.22
5.4	3.9	Setosa	0.82
7.4	2.8	Virginica	2.22
6.1	2.8	Versicolor	0.94
7.3	2.9	Virginica	2.1
6.0	2.7	Versicolor	0.89
5.8	2.8	Virginica	0.67
6.3	2.3	Versicolor	1.36
5.1	2.5	Versicolor	0.60
6.3	2.5	Versicolor	1.25
5.5	2.4	Versicolor	0.75

Sepal Length	Sepal Width	Species
5.2	3.1	?

يمكن استخدام خوارزمية KNN في توصية أشياء أخرى عن طريق الاستنتاج من تفضيلات المستخدمين السابقة وتوصية العناصر الجديدة بناءً على الجيرة الأقرب تطبق على مجموعة متنوعة من المجالات. هنا بعض الأمثلة:

1. توصية الموسيقى: يمكن استخدام **KNN** لتوصية الأغاني أو الفنانين للمستخدمين بناءً على تفضيلاتهم الموسيقية السابقة وتشابه التذاكر الصوتية بين المستخدمين.

2. توصية المنتجات التجارية: يمكن استخدام **KNN** لتوصية المنتجات المشابهة أو المنتجات المرتبطة للمستخدمين بناءً على تاريخ مشترياتهم السابقة وتفضيلاتهم.

إن استخدام KNN في توصية أشياء أخرى يعتمد على توفر بيانات التفضيلات ومعلومات الجيرة المناسبة. يجب تحديد السياق والمجال الذي ترغب في توصية العناصر فيه وتكوين مجموعة بيانات مناسبة قبل تطبيق الخوارزمية KNN.