

K-En Yakın Komşular

K-En Yakın Komşular, Makine Öğrenimi'nde en temel ve önemli sınıflandırma algoritmalarından biridir. Denetimli öğrenme alanına aittir ve desen tanıma, veri madenciliği ve saldırı tespiti gibi alanlarda yoğun uygulama bulur. K-En Yakın Komşular (KNN) algoritması, hem sınıflandırma hem de regresyon problemlerini çözmek için kullanılabilen basit, uygulanması kolay bir denetimli makine öğrenme algoritmasıdır. KNN algoritması, benzer şeylerin yakın çevrede bulunduğunu varsayar. Başka bir deyişle, benzer şeyler birbirine yakındır. KNN, çocukluğumuzda öğrenmiş olabileceğimiz bazı matematiksel yöntemlerle (bir grafikteki noktalar arasındaki mesafeyi hesaplama) benzerlik kavramını (bazen mesafe, yakınlık veya yakınlık olarak da adlandırılır) yakalar. Çözmeye çalıştığımız probleme bağlı olarak tercih edilebilecek başka mesafe hesaplama yöntemleri de vardır. Ancak, doğru çizgi mesafesi (aynı zamanda Öklid mesafesi olarak da adlandırılır) popüler ve tanıdık bir seçenektir. Parametrik olmayan bir yöntem olduğundan, gerçek hayattaki senaryolarda geniş çapta kullanılabilir, yani, veri dağılımı hakkında herhangi bir temel varsayımda bulunmaz (GMM gibi diğer algoritmaların aksine, verilen verinin Gaussian dağılımını varsayarlar). Bu makale, sklearn kütüphanesini kullanarak örnek rastgele veriler üzerinde K-en yakın komşuları göstermektedir.

Uygulama

Prostat kanseri verisi ile yapılan model eğitimi tahmin sonuçlarının k sayısının değişimine göre model başarısının değişimi aşağıdaki gibi olmuştur.

k'nın 1 değeri için					
[[22 1]					
[3 4]]					
	precision	recall	f1-score	support	
0	0.88	0.96	0.92	23	
1	0.80	0.57	0.67	7	
accuracy			0.87	30	
macro avg	0.84	0.76	0.79	30	
weighted avg	0.86	0.87	0.86	30	
k'nın 3 değeri için					
[[21 2]					
[4 3]]					
	precision	recall	f1-score	support	
0	0.84	0.91	0.88	23	
1	0.60	0.43	0.50	7	
accuracy			0.80	30	
macro avg	0.72	0.67	0.69	30	
weighted avg	0.78	0.80	0.79	30	
k'nın 5 değeri için					
[[21 2]					
[4 3]]					
	precision	recall	f1-score	support	
0	0.84	0.91	0.88	23	
1	0.60	0.43	0.50	7	
accuracy			0.80	30	
macro avg	0.72	0.67	0.69	30	
weighted avg	0.78	0.80	0.79	30	

k'nın 7 değeri için					
[[22 1]					
[4 3]]					
	precision	recall	f1-score	support	
0	0.85	0.96	0.90	23	
1	0.75	0.43	0.55	7	
accuracy			0.83	30	
macro avg	0.80	0.69	0.72	30	
weighted avg	0.82	0.83	0.82	30	
k'nın 9 değeri için					
[[22 1]					
[4 3]]					
	precision	recall	f1-score	support	
0	0.85	0.96	0.90	23	
1	0.75	0.43	0.55	7	
accuracy			0.83	30	
macro avg	0.80	0.69	0.72	30	
weighted avg	0.82	0.83	0.82	30	