

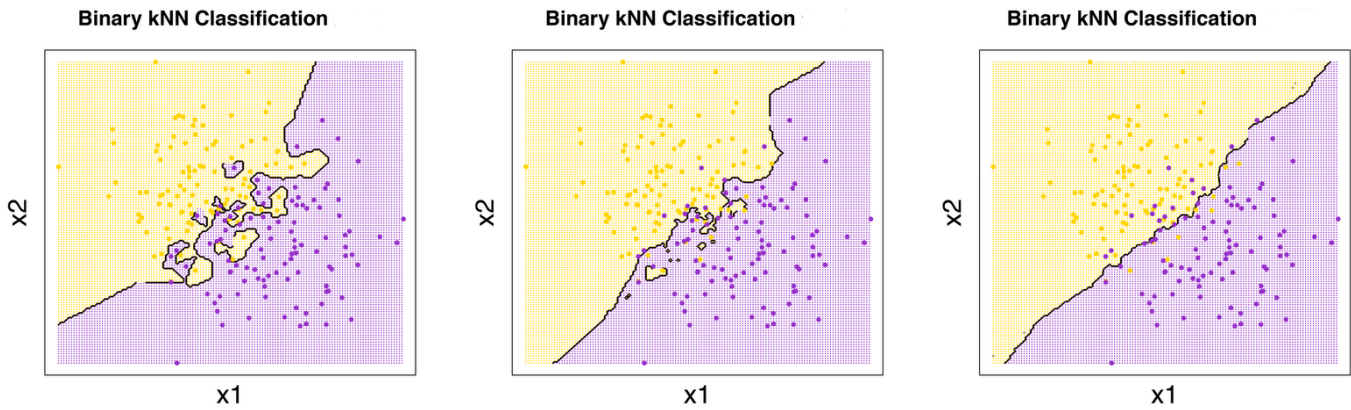
Ödev #2

1 Ön bilgiler

KDDCUP'99 saldırı tespit sistemleri için üretilen veri kümesiyle kNN ve Kümeleme algoritmalarının kullanımı.

2 Sorular

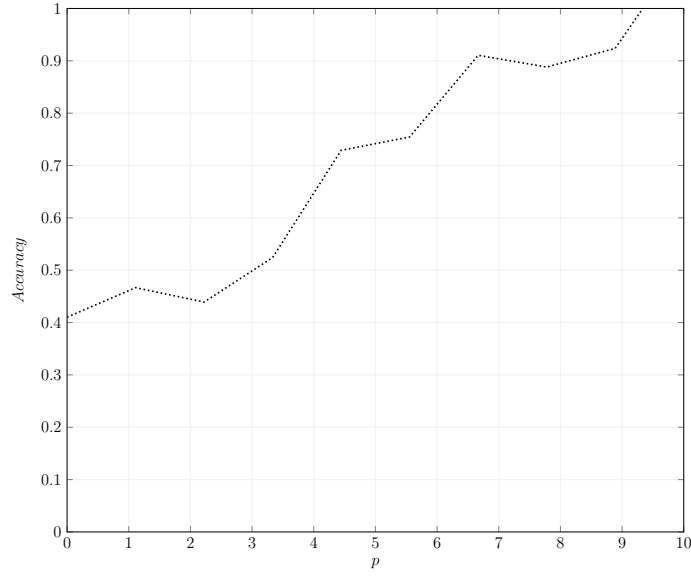
1. KDDCUP'99 veri kümesinden derste lab uygulamasında kullanmadığımız 2 adet nitelik kullanarak $k = \{1, 3, 5\}$ değerleri için kNN karar sınırlarını 3 farklı grafikte gösteriniz. Matplotlib kütüphanesi içinde subplot metodunu kullanarak grafikleri tek satırda yanyana gösterebilirsiniz. Örnek grafik:



2. KDDCUP'99 verikümesinde bulunan bütün sayısal alanları kullanarak, kNN algoritması için $k = \{1, 3, 5, 7, 10\}$ olarak seçtiğinizde *sklearn.metrics.accuracy_score* metodunu kullanarak oluşan sınıflandırma modellerinin, sınıflandırma performansını gösteriniz. En iyi sınıflandırma performansını hangi k değeri ile elde ettiniz?

1	3	5	7	10

3. kNN için $k = 10$ ve Minkowski metrikleri ($p = \{1, 2, \dots, 10\}$) olacak şekilde sınıflandırma modelinin doğruluk oranının (*sklearn.metrics.accuracy_score*) değişimini grafikte gösteriniz. Örnek grafik (sadece x ve y axislerinde yer alan bilgileri göstermektedir. Sonuç bu grafikte olduğu gibi doğrusal bir değişim göstermeyebilir.)



4. Scikit-learn kütüphanesinde yer alan *make_blobs* metodunu kullanarak $n_samples=30000$, $n_features=30$, $centers=5$ parametreleriyle sentetik veri kümesi oluşturulacaktır. k değeri 5 olacak şekilde veri kümesinde k-Means, Hierarchical clustering (*SciPy* kullanmayın, *sklearn* kullanın. Kullanılması gereken sınıf: *sklearn.cluster.AgglomerativeClustering*), DBSCAN kümeleme algoritmaları için *homogeneity*, *completeness* ve *silhouette* (*sklearn.metrics.silhouette_score*) değerlerini bir tabloda gösteriniz. Örnek tablo

Table 1: Örnek tablo.

Algoritma	Homogeneity	Completeness	Silhouette
K-means			
Hierarchical			
DBSCAN			

5. k-means için $k = \{2, \dots, 10\}$ olacak şekilde KDDCUP'99 veri kümesi kullanarak her bir k değeri için Silhouette hesaplayınız. En ideal k değerini bulunuz. *Pandas* kütüphanesini kullanarak veri kümesinden 30.000 adet örneklem alarak çözünüz.

Table 2: Örnek tablo

	2	3	4	5	6	7	8	9	10
Silhouette									

3 Notlandırma

Soruların çözümünü **Jupyter Notebook (*.ipynb)** dosyasını çalıştırarak çıktıları dosyada görünecek şekilde ozgur.catak@tubitak.gov.tr e-posta adresine gönderiniz.

Ödev son teslim tarihi 08.04.2018 23:59'dur. Bu tarihten sonra her gün için 25 puan kırılabacaktır.