

2023-2024 Güz Yarıyılı Makine Öğrenmesi Dersi 2. Ödevi

Konu: Bu ödevde verilen resimleri k-means yöntemi ile renk benzerliklerine göre kümeleyen bir sistem tasarmanız ve gerçeklemeniz istenmektedir.

Veri Kümesi: Programın çalışma başarısını <https://universe.roboflow.com/example01/color-oquyt> linkinden erişebileceğiniz veri kümesindeki red, blue, green, gray, white sınıflarında bulunan resimleri kullanarak değerlendiriniz. Her sınıf için ilk 20 resmi kullanınız.

Yapılacak İşlemler: Ödevde aşağıdaki işlemleri yapınız:

1. Resimlerin formatını hazır kütüphaneler kullanarak çözerek görüntüye ait matrisi elde ediniz. Bir görüntü piksellerin (R,G,B) bileşenlerinden oluştuğu bir matristir.
2. Resimleri oluşturan piksellerin (R,G,B) değerlerine göre her resmin renk histogramını (Her resimde R, G, B bileşenleri için ayrı ayrı olmak üzere toplam 3 histogram) elde ediniz. Histogram eldesi için de hazır kütüphane kullanabilirsiniz. Bir pikselin R,G,B bileşenlerinin değeri 0-255 arası değiştiği için her histogram dizisi 256 elemanlı olmalıdır.
3. Her resimde, her renk bileşeni için histogram dizisindeki sonuçları resimdeki toplam piksel sayısına bölerek dizi elemanlarının değerini [0-1] aralığına normalize ediniz.
4. Resimlerin histogramlarını benzerliklerine göre $k=5$ değeri için (5 renk sınıfı olduğu için) k-means yöntemiyle kümeleyiniz. K-means'de başlangıç adımında rasgele seçeceğiniz k resmin histogramını k cluster'ın başlangıç merkezi olarak kullanınız.

Ödev Kodlaması: Kodlama için python kullanınız. **K-Means** algoritmasının kodunu kendiniz yazınız. Diğer işlemlerin tamamı için hazır kütüphanelerden yararlanabilirsiniz. Kütüphane olarak Numpy, Pandas, scikit-learn, openCV ve görseller için matplotlib kullanabilirsiniz.

Ödev Raporu: Ödev raporunuzu aşağıdaki sıra ile **makale formatında** yazınız.

1. **Özet:** Yaptığınız çalışmayı, elde ettiğiniz sonuçları özet olarak veriniz.
2. **Giriş:** Ödev konusunu tanıtan 1 paragraflık bir giriş yapınız. Bu çalışmanın nerelerde kullanılabileceğinden bahsediniz.
3. **Sistem Tasarımı:** Sisteminizin işlem adımlarını kısaca anlatınız.
4. **DeneySEL Sonuçlar:** Sistem başarısını değerlendirmek için aşağıdaki işlemleri yapınız.
 - a. Küme merkezlerini rasgele belirleyerek her k değeri için kümeleme işlemini 10 defa tekrarlayınız. Her kümeleme sonunda, kümelerde doğru cluster'da olan resim yüzdesini hesaplayınız.
 - b. Hazırlayacağınız karışıklık matrisi (confusion matrix) üzerinde sonucu gösteriniz.
 - c. Her sınıf için doğru cluster'a yerleşmiş 5 örnek ve yanlış cluster'da bulunan 1 örnek resmi veriniz.
5. **Sonuç:** Sizce kümeleme işlemi başarılı oldu mu? Yanlış kümelerde olan resimler sizce neden yanlış kümelerde yer aldı? Sistemin genel başarısını yorumlayınız. Başarısızlık sebebi olduğunu düşündüğünüz problemlerin giderilmesi için varsa önerilerinizi belirtiniz.

Ödev Teslimi: Ödev raporunuzu (pdf olarak) ve python'da yazdığınız program kodunuzu birlikte zip'leyerek sisteme yükleyiniz. Rapor **%30**, kod **%70** olarak değerlendirilecektir.

Ödev Teslim Tarihi: Ödevinizi **7 Ocak 2024 23.59'a** kadar sisteme yükleyiniz.

Değerlendirme: Ödevlerinizi yaparken internetteki kaynaklardan yararlanabilir, arkadaşlarınız ile fikir alışverişi yapabilirsiniz. Fakat internetteki bir koda veya bir başka ödevle **%70'den fazla** benzeyen ödevler kopya olarak değerlendirilir.