

## İKT484 Bahar 2025

### Ödev 3

**Amaç:** Boyut küçültmenin analizi, tasarımı ve uygulanması konusunda deneyim kazanmak.

#### Açıklama:

Sınıf içi Lineer Ayrım Analizi (LDA) örneğini ele alalım. Bu örnekte

- TFF Süper Lig'in 2022-2023 ve 2023-2024 sezonlarının 34. haftasından etiketli veriler. Etiketler biraz keyfi olarak yapılmıştı.
- Aynı ligin 2024-2025 sezonunun 34. haftasından etiketsiz veriler. Bunlar için etiketler LDA tarafından atandı.

Şimdi sizden bazı değişiklikler yapmanız, sonuçları gözlemlemeniz ve yorum yapmanız istenecektir.

1. Aşağıdaki özelliklere sahip benzer bir veri kümesi oluşturun
  1. COVID molasını atlayarak 5 sezon için aynı ligin 17. haftasından verileri alın (2022-2023, 2021-2022, 2018-2019, 2017-2018, 2016-2017).
  2. Her sezonun son haftasını kullanarak takımları etiketleyin, böylece etiketler tamamen nihai sonuçlar (yani son puan durumu) olsun.
    1. Şampiyon (Tek Takım, 1.lik)
    2. İkinci (Tek Takım, 2.lik)
    3. Avrupa (Birden fazla takım, Avrupa kupalarına katılmaya hak kazananlar)
    4. Orta (Diğer tüm takımlar)
    5. Alt (En alttaki 3 veya 4 takım, küme düşmeye tabi)
  3. Geçtiğimiz yılın sezonunun (2023-2024) 17. hafta verilerini etiketsiz ekleyin. O sezon için kimin şampiyon olduğunu vb. bildiğimizi unutmayın.
2. Aynı algoritmayı çalıştırın (kod için çok fazla değişiklik gerekmiyor, ancak istediğiniz gibi yapın).
3. Sonuçları, özellikle yanlış sınıflandırılmış (yani yanlış) tahminlere odaklanarak tartışın.
  1. Sınıflandırma şeması Şampiyon ve İkinci'yi doğru şekilde sınıflandırıyor mu? Sonuçları, "dengesiz veya seyrek sınıf dağılımları" kavramı ve bunun özellikle LDA olmak üzere sınıflandırma algoritmalarını nasıl etkilediği hakkındaki araştırmanızla tartışın.
  2. Tüm sınıflar için algoritma sonucunu ve manuel sınıflandırma sonuçlarını karşılaştırarak Gerçek Pozitif, Gerçek Negatif, Yanlış Pozitif, Yanlış Negatif değerlerini hesaplayın.

3. LDA'nın bir tahmin yöntemi olarak bu uygulaması ne kadar güvenilirdir?
4. Sezon ortası istatistiklerine göre takım sonuçlarını sınıflandırmak için LDA'yı kullanmanın ima ettiği sıfır ve alternatif hipotezleri belirtin. Örneğin: 'H0: Sezon ortası takım performansı sezon sonu sınıflandırmasını önemli ölçüde tahmin etmez.'
5. Şampiyon, İkinci, Orta vb. etiketlerini bir taksonominin veya ontolojinin parçası olarak mı değerlendirirsiniz? Ontolojik bir yapı bu bilgiyi temsil etmede nasıl farklılık gösterir?
6. Aynı veriler üzerinde bir kümeleme algoritması (örneğin k-means) çalıştırmayı deneyin ve gruplamaların denetlenen etiketlerden nasıl farklılaştığını karşılaştırın. Bu size ontoloji ile veri odaklı grup ortaya çıkışı hakkında ne anlatıyor?