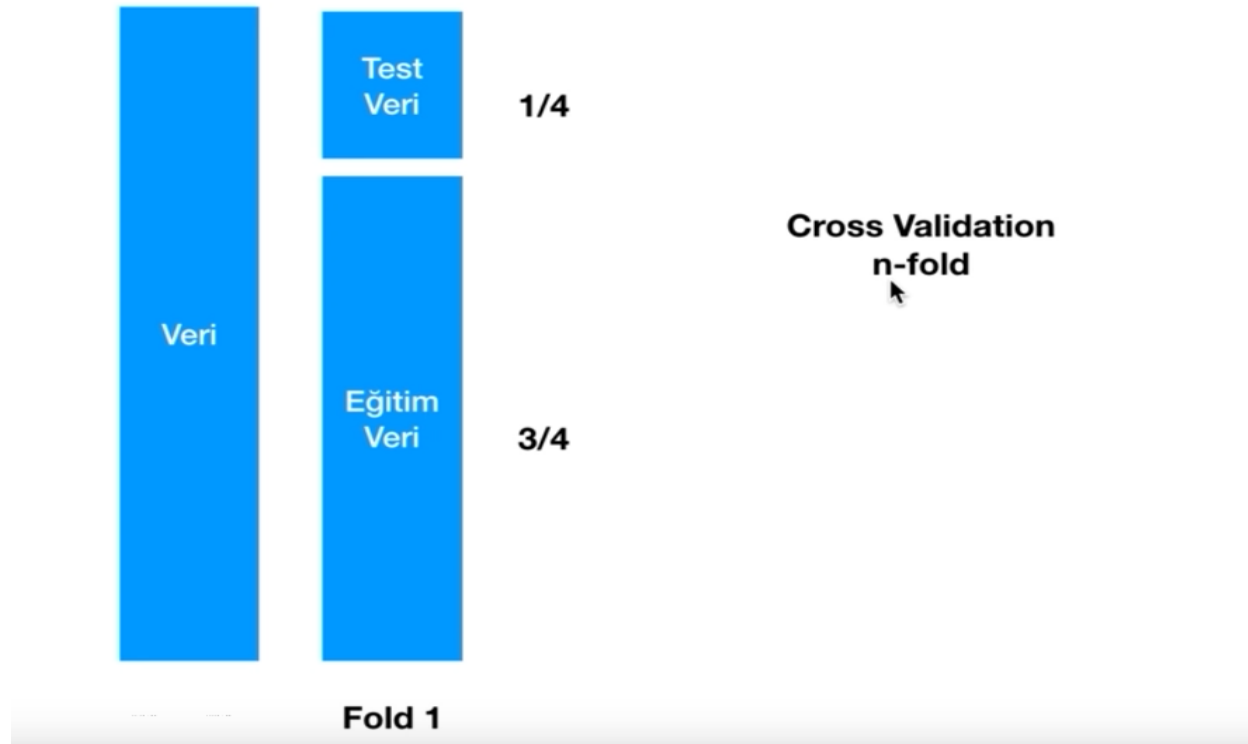


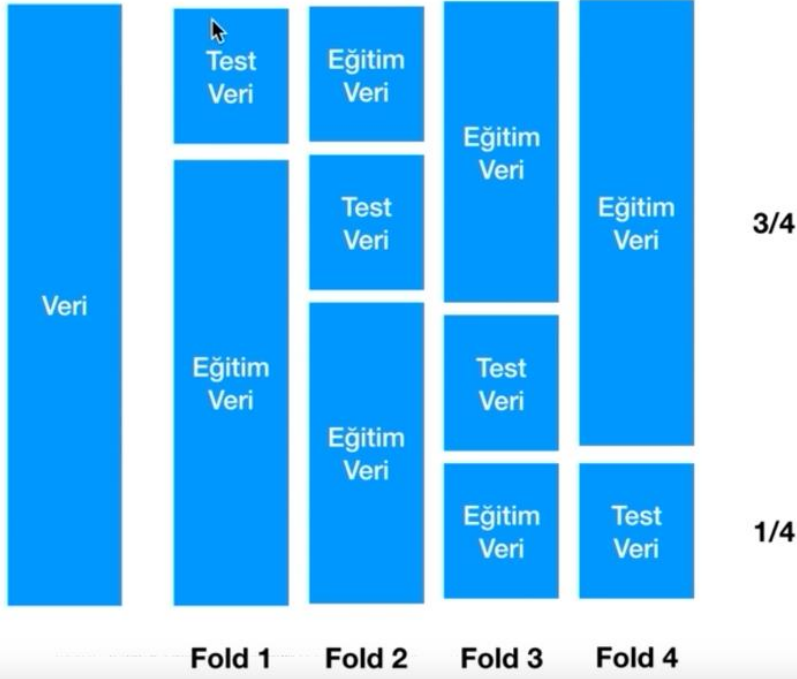
Modellerin Değerlendirilmesi ve Seçimi

- Modellerin başarısı, parametrelere bağlıdır
- Makine öğrenmesi parametreleri optimize etmez
- Model seçiminden önce dikkat edilmesi gereken ilk nokta, model değerlendirilmesidir (evaluation)
- Şimdiye kadar test kümesindeki başarıyı ölçtük
- k-fold Cross Validation (k-katlamalı Çapraz Doğrulama)
- Izgara Araması (Grid Search)

k-fold cross validation: k= 4 için



Cross Validation n-fold



Çapraz doğrulamadaki amaç bütün veri setini eğitim ve test olarak değerlendirebilmek, önceki model başarı değerlendirmelerinde sadece test verisi üzerindeki başarıya göre yapıyoduk ve bu büyük resmi görmemize engel, fakat k-fold cross validation bütün veriyi kullanarak daha kullanışlı bir değerlendirme yapıyor.

Hangi Modeli Seçeceğim?

- Öncelikle Modelin Tipinin Belirlenmesi
 - Bağımlı Değişken Var mı? (Classification-Regression / Clustering)
 - Bağımlı Değişken Kategorik mi sürekli bir sayı mı? (Classification / regression)
 - Doğrusal / doğrusal olmayan

Parametreleri optimize etmek, bir makine öğrenimi modelinin performansını artırmak için kullanılan bir yöntemdir. Bir modelin performansını etkileyen birçok parametre veya hiperparametre vardır. Örneğin, destek vektör makineleri için C değeri, karar ağaçları için maksimum derinlik, k-NN için komşu sayısı gibi.

Parametre optimizasyonu, bu parametrelerin en iyi değerlerini bulmayı amaçlar. Bu, genellikle deneme-yanılma yöntemiyle yapılır, yani farklı parametre değerlerini deneyerek ve modelin performansını ölçerek gerçekleştirilir. Genellikle çapraz doğrulama gibi teknikler kullanılarak, her bir parametre ayarının modelin genel performansını nasıl etkilediği değerlendirilir. Birkaç parametre kombinasyonu denendiğinde, en iyi sonuçları veren parametre değerleri seçilir ve bu değerlerle model tekrar eğitilir. Bu şekilde, modelin en iyi performansı elde etmesi sağlanır.

Parametre optimizasyonu, genellikle hiperparametre araması (hyperparameter search) veya hiperparametre ayarlama (hyperparameter tuning) olarak da adlandırılır. Bu süreç, modelin daha iyi performans göstermesini sağlayarak, genellikle daha doğru ve güvenilir tahminler elde etmeye yardımcı olur.

Grid Search, bir makine öğrenimi modelinin performansını optimize etmek için kullanılan bir hiperparametre optimizasyon tekniğidir. Makine öğrenimi modelleri, performanslarını etkileyen bir dizi hiperparametreye sahiptir. Örneğin, destek vektör makineleri için C değeri, karar ağaçları için maksimum derinlik, k-NN için komşu sayısı gibi.

Grid Search, modelin performansını en iyi hale getirmek için bu hiperparametrelerin en iyi kombinasyonunu bulmayı amaçlar. Bu, belirli bir hiperparametre alanındaki olası değerlerin bir ızgara (grid) üzerinde düzenlenmesi ve her bir kombinasyonun modelin performansı ile değerlendirilmesiyle yapılır.

Grid Search işlemi şu adımları içerir:

Hiperparametre alanlarının belirlenmesi: Grid Search yapılacak hiperparametrelerin ve bu hiperparametrelerin alabileceği değerlerin belirlenmesi.

Grid oluşturma: Belirlenen hiperparametre alanlarındaki değerlerin tüm kombinasyonlarını içeren bir ızgara oluşturulur.

Her bir kombinasyonun model performansının ölçülmesi: Her bir ızgara noktası için, belirlenen hiperparametrelerle model eğitilir ve çapraz doğrulama gibi bir değerlendirme yöntemi kullanılarak performansı ölçülür.

En iyi performansı veren hiperparametre kombinasyonunun seçilmesi: Grid Search, en iyi performansı veren hiperparametre kombinasyonunu belirler.