

## 2.2 Büyük Resmi Görmek

Bu bölümde bazı sorulara yanıt aramanız beklenmektedir. Bazen elinizde problem vardır bir çözüm ararsınız bazen ise elinizdeki veri seti ile neler yapabileceğinize karar verirsiniz. Bu aşamada büyük resmi görmek önemlidir. Büyük resimden kasıt sistemin ana hatlarını oluşturabilmektedir. Sistemin ana hatlarını oluştururken aşağıdaki sorulara yanıtlar bulmak işinizi kolaylaştıracaktır. Böylece ileriki aşamalarda, sistemi baştan kurmak zorunda kalmazsınız.

1. Problemi açık bir şekilde tanımlayın.
2. Problem hangisine ait?
  1. Değer Tahmin Etmek
  2. Alışılmadık Olayı Bulmak
  3. Örüntü Keşfetmek
  4. Kategori Tahmin Etmek
3. Problemin çözümünün nasıl kullanılacağını karar verin.
4. Bu problem için kullanılan başka çözümler var mıdır?
5. Sistem nasıl planlanmalıdır?
  1. Denetimli - Denetimsiz Öğrenme
  2. Toplu - Çevrimiçi Öğrenme
6. Sistemin ana hatları nasıl olacaktır?
  1. Veri nereden okunacak
  2. Nerede depolanacak
  3. Sonuçlar nereye kaydedilecek
7. Sistemin performansı nasıl ölçülmelidir?
8. Problemi elle nasıl çözerdiniz?
9. Varsa probleminizin çözümü ile ilgili varsayımları listeleyin.

### 2.2.1 Sorunun Hatlarını Çizmek

Makine öğrenmesi ile amaçladığınız şeyin ne olduğuna iyi karar vermelisiniz. Net olmayan problemlere rastgele çözümler aramak yerine, cevabına ulaşmak istediğiniz sorunun hatlarını doğru çizmek ileride uygulayacağınız makine öğrenmesi algoritmasını doğru seçmenizi sağlayacaktır.

Örneğimizle devam edecek olursak elimizde iki tane çözümünü öğrenmek istediğimiz sorun var. Daha önce belirtildiği gibi bu bölümün amacı sadece herhangi bir makine

öğrenmesi uygulaması yapılırken nasıl bir yol izlenmesi gerektiği ile ilgili bir fikir verebilmektir.

1. Ödenen hesaba göre verilen bahşışler nasıl değışir?
2. 26.5 lira hesap ödeyen bir kadın ne kadar bahşış bırakır?

Sorunu değışkenler arasındaki ilişkiyi ölçerek bir sonraki değeri tahmin etmek diye tanımlayabiliriz. Burada değışkenler; ödenen hesap ve verilen bahşış miktarıdır. Girdi değışkenlerini bazı sürekli fonksiyonlara eşlemeye çalışacağız, bu bir regresyon problemidir. Denetimli ve toplu öğrenme kullanacağız. Daha karmaşık problemlerde yukarıda verilen soruların hepsini cevaplamalıyız, ancak şimdi başka çözümler var mıdır, elle nasıl çözeriz, sistemin ana hatlarını nasıl tanımlamalıyız gibi soruları yanıtlamayacağız.

### 2.2.2 Performans Ölçütünü Seçmek

Bu aşamada makine öğrenme algoritmanızın ne kadar doğru çalıştığını anlayabilmek için bir performans ölçütü seçmelisiniz. Sisteminizin doğruluğunu ölçmek için kullanacağınız metrik bütün makine öğrenmesi algoritmalarına uygun olmayabilir. Bu nedenle regresyon probleminde kullanacağınız metriği herhangi bir sınıflandırma algoritmasında kullanmaya çalışırsanız her zaman doğru olmayabilir.

Regresyon problemlerinde sıklıkla kullanılan bir metrik olan Kare Ortalama Hata (Mean Squared Error) fonksiyonunu sistemimizin performansını ölçmek için kullanacağız.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{y}_i)^2$$