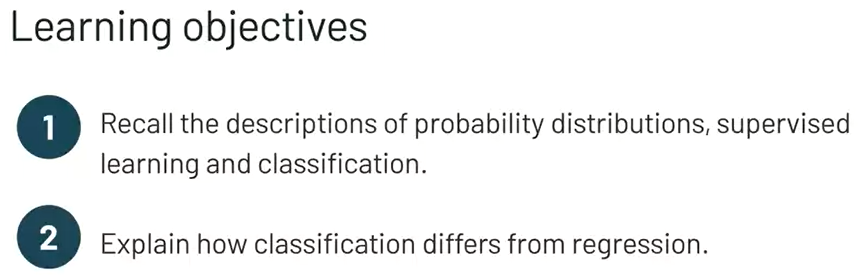
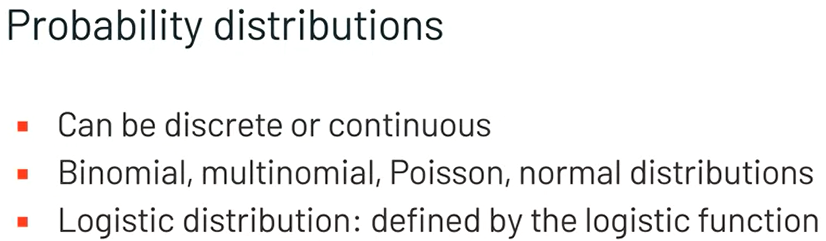
Classification Modelling

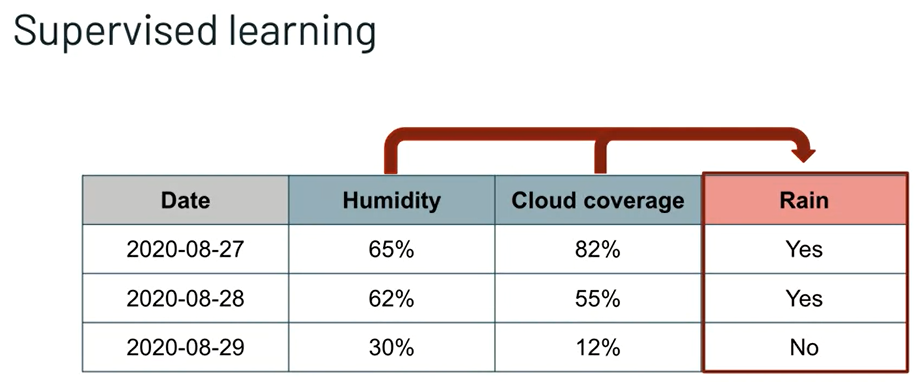
A Review of Key Concepts



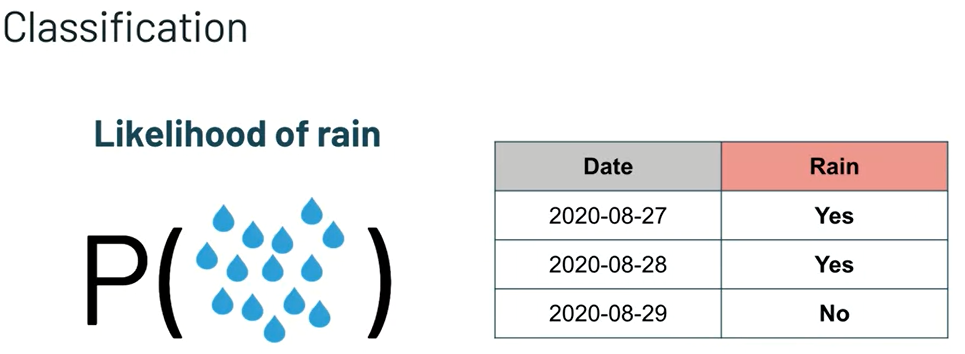
Welcome to Lesson 3, Classification Modeling. [Ders 3, Sınıflandırma Modelleme'ye hoş geldiniz.] In this lesson, you'll learn how to use and evaluate logistic regression for classification problems. [Bu derste, sınıflandırma problemleri için lojistik regresyonu nasıl kullanacağınızı ve değerlendireceğinizi öğreneceksiniz.] You will also learn some of the key concepts associated with classification modeling. [Ayrıca sınıflandırma modellemesi ile ilgili bazı temel kavramları da öğreneceksiniz.]



Previously in this course, we learned about some of the different kinds of probability distributions, for both discrete and continuous variables. [Bu kursun önceki bölümlerinde, hem kesikli hem de sürekli değişkenler için farklı olasılık dağılımlarından bazılarını öğrenmiştik.] Some of these were binomial, multinomial, Poisson, Bernoulli, and normal distributions. [Bunlardan bazıları binom, çok terimli, Poisson, Bernoulli ve normal dağılımlardı.] In addition, we also talked about a logistic distribution. [Ayrıca lojistik dağıtımdan da bahsettik.] The shape of this distribution is defined by the logistic function, it looks similar to the normal distribution, in that it's symmetric and has one peak, but it has a longer tail on one side. [Bu dağılımın şekli lojistik fonksiyon tarafından tanımlanır, simetrik olması ve bir tepe noktasına sahip olması, ancak bir tarafında daha uzun bir kuyruğu olması bakımından normal dağılıma benzer.] It's frequently used to model growth and for things like sports, physical sciences, finance, and other topics. [Büyümeyi modellemek ve spor, fizik bilimleri, finans ve diğer konular gibi şeyler için sıklıkla kullanılır.] It's a better fit than the normal distribution for extreme events or outliers. [Aşırı olaylar veya aykırı değerler için normal dağılımdan daha uygundur.]



We also learned previously about the distinction between supervised and unsupervised machine learning problems. [Denetimli ve denetimsiz makine öğrenimi sorunları arasındaki farkı da daha önce öğrenmiştik.] Supervised learning is when an algorithm is learning a function that maps a collection of inputs to an output based on example pairs of inputs and output, which is our training data. [Denetimli öğrenme, bir algoritmanın, eğitim verilerimiz olan örnek girdi ve çıktı çiftlerine dayalı olarak bir girdi koleksiyonunu bir çıktıya eşleyen bir işlevi öğrenmesidir.] In a way, our algorithm is being supervised by certain truth values for the function's output. [Bir bakıma algoritmamız, fonksiyonun çıktısı için belirli doğruluk değerleri tarafından denetleniyor.] These truth values, also called labels, are the key behind supervised learning. [Etiketler olarak da adlandırılan bu doğruluk değerleri, denetimli öğrenmenin ardındaki anahtardır.] We can't do supervised learning without these supervising labels. [Bu denetleyici etiketler olmadan denetimli öğrenme yapamayız.] This is in contrast to unsupervised learning, where algorithms look to understand relationships between inputs without any preexisting labels. [Bu, algoritmaların önceden var olan herhangi bir etiket olmadan girdiler arasındaki ilişkileri anlamaya çalıştığı denetimsiz öğrenmenin aksine.] In this module, we'll be focusing on classification, which is the type of machine learning tasks, within supervised learning. [Bu modülde, denetimli öğrenme kapsamında makine öğrenimi görevlerinin türü olan sınıflandırmaya odaklanacağız.]

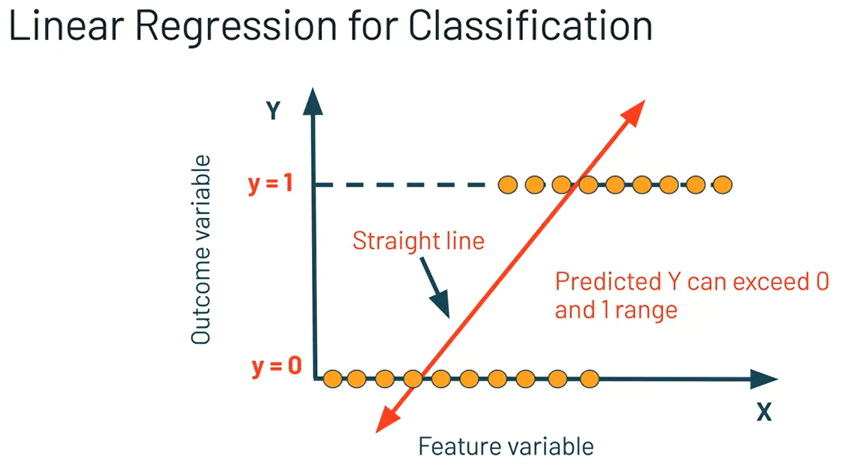


We previously defined classification as a subset of supervised learning, where the model is trying to predict which class out of a pre-defined set of classes a record belongs to. [Sınıflandırmayı daha önce, modelin bir kaydın önceden tanımlanmış bir sınıflar kümesinden hangi sınıfa ait olduğunu tahmin etmeye çalıştığı, denetimli öğrenmenin bir alt kümesi olarak tanımlamıştık.] Now that we've reviewed the descriptions of probability distributions, supervised learning, and classification, we'll move on to the next video, where we'll talk about logistic regression used as a classification modeling algorithm. [Olasılık dağılımlarının, denetimli öğrenmenin ve sınıflandırmanın açıklamalarını gözden geçirdiğimize göre, sınıflandırma modelleme algoritması olarak kullanılan lojistik regresyon hakkında konuşacağımız sonraki videoya geçeceğiz.]

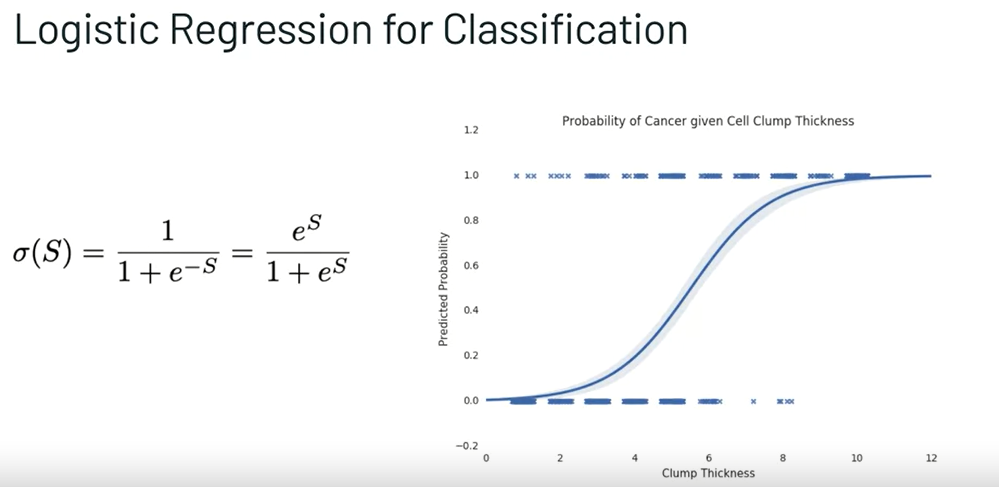
# An Introduction to Logistic Regression

# 

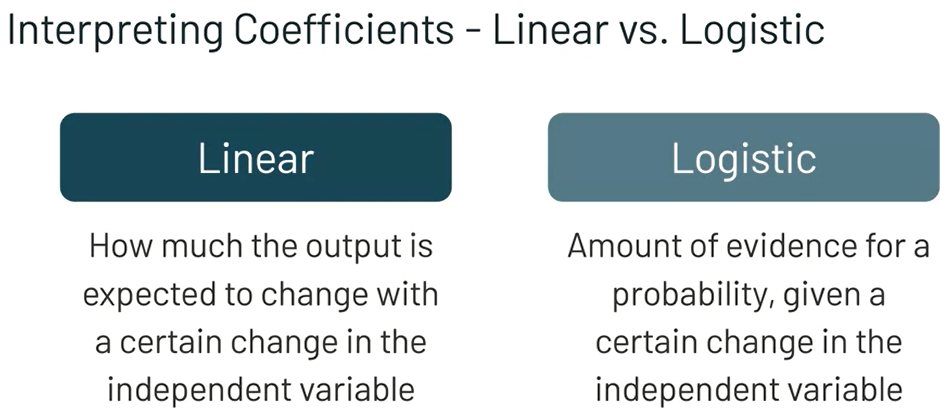
Within the field of machine learning we have several types of algorithms available to perform classification tasks. [Makine öğrenimi alanında, sınıflandırma görevlerini gerçekleştirmek için kullanılabilecek çeşitli algoritma türlerine sahibiz.]



One of the most commonly used is called logistic regression, which is what will be learning about in this video. [En yaygın olarak kullanılanlardan biri, bu videoda öğrenilecek olan lojistik regresyondur.] Previously we talked about some typical uses for Linear Regression, however, it's not well suited for classification tasks. [Daha önce Lineer Regresyonun bazı tipik kullanımlarından bahsetmiştik, ancak sınıflandırma görevleri için pek uygun değil.] This is because Linear Regression predicts continuous, not probabilistic outputs. [Bunun nedeni, Doğrusal Regresyon'un olası çıktıları değil, sürekliliği öngörmesidir.] In a binary classification problem, what we're interested in is the probability of an outcome occurring. [İkili bir sınıflandırma probleminde ilgilendiğimiz şey, bir sonucun meydana gelme olasılığıdır.] Probability is ranged between 0 and 1, where the probability of something that's definitely certain to happen is 1 and something unlikely to happen or definitely won't happen is 0. [Olasılık 0 ile 1 arasında değişir, burada kesinlikle olması kesin olan bir şeyin olasılığı 1 ve olması muhtemel olmayan veya kesinlikle olmayacak bir şeyin olasılığı 0'dır.] But in Linear Regression, were predicting an absolute number which can range outside 0 and 1 up to potentially infinity. [Ancak Lineer Regresyonda, 0 ve 1'in dışında potansiyel olarak sonsuzluğa kadar değişebilen mutlak bir sayı tahmin ediyordu.] If we try to draw a straight line to fit a function to the data, we see that it doesn't fit this binary probabilistic outcome very well at all. [Verilere bir fonksiyon sığdırmak için düz bir çizgi çizmeye çalışırsak, bu ikili olasılık sonucuna pek de uymadığını görürüz.] The function can predict a negative probability or a probability greater than 1, which would be a less than 0 or more than 100% chance of a given outcome. [İşlev, belirli bir sonucun 0'dan küçük veya %100'den fazla şansı olan negatif bir olasılığı veya 1'den büyük bir olasılığı tahmin edebilir.] That's obviously not realistic and doesn't make any sense. [Bu açıkça gerçekçi değil ve hiçbir anlam ifade etmiyor.] Another reason that Linear Regression doesn't perform classification well is that it doesn't work well with a very unbalanced data set. [Doğrusal Regresyon'un sınıflandırmayı iyi yapmamasının bir başka nedeni de, çok dengesiz bir veri seti ile iyi çalışmamasıdır.] Where there's much more of one class represented in the training data than another class. [Eğitim verilerinde başka bir sınıftan çok daha fazla bir sınıfın temsil edildiği yer.]

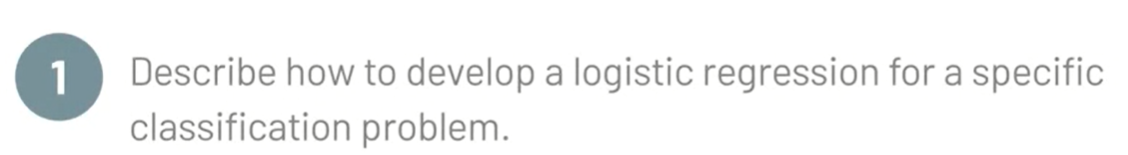


The Logistic function, which is what Logistic Regression uses, draws an S or a sigmoid curve to fit the data and will predict a probability between 0 and 1 as an output of the model. [Lojistik Regresyonun kullandığı Lojistik işlevi, verilere uyması için bir S veya sigmoid eğrisi çizer ve modelin çıktısı olarak 0 ile 1 arasında bir olasılık tahmin eder.] Here we see an example which is from a famous data set about Cancer and tumors. [Burada Kanser ve tümörler hakkında ünlü bir veri setinden bir örnek görüyoruz.] Where they probability of Cancer we see on the y-axis between 0 and 1, an Clump Thickness is one particular variable in the data set. [Y ekseninde 0 ile 1 arasında gördüğümüz Yengeç olasılığının olduğu yerde, bir Küme Kalınlığı veri setindeki belirli bir değişkendir.] And the equation on the left is just the sigmoid equation. [Ve soldaki denklem sadece sigmoid denklemdir.]



Interpreting the Coefficients of Logistic regression is trickier than it is for Linear regression. [Lojistik regresyonun Katsayılarını yorumlamak, Lineer regresyon için olduğundan daha zordur.] In Linear regression, the coefficients tell us if variable x goes up by a certain amount, how would we expect the value of the output to change? [Doğrusal regresyonda, katsayılar bize x değişkeninin belirli bir miktarda artması durumunda çıktının değerinin nasıl değişmesini bekleyeceğimizi söyler?] In Logistic regression, we can interpret a coefficient as the amount of evidence provided per change in the associated predictor variable or x, rather than a set amount. [Lojistik regresyonda, bir katsayıyı, belirli bir miktar yerine ilişkili tahmin değişkeni veya x'teki değişiklik başına sağlanan kanıt miktarı olarak yorumlayabiliriz.] We also say that the coefficient is the log odds of a particular outcome. [Katsayının belirli bir sonucun log oranları olduğunu da söylüyoruz.] It's important to also note that this coefficient is not constant across all values of the feature variable. [Bu katsayının, özellik değişkeninin tüm değerlerinde sabit olmadığına da dikkat etmek önemlidir.] Now that we've introduced some of the concepts of Logistic regression and how it differs from Linear regression in the next video, will talk about how to actually apply Logistic regression for a particular classification problem. [Şimdi, bir sonraki videoda Lojistik regresyonun bazı kavramlarını ve Lineer regresyondan nasıl farklı olduğunu tanıttığımıza göre, belirli bir sınıflandırma problemi için Lojistik regresyonun gerçekte nasıl uygulanacağı hakkında konuşacağız.]

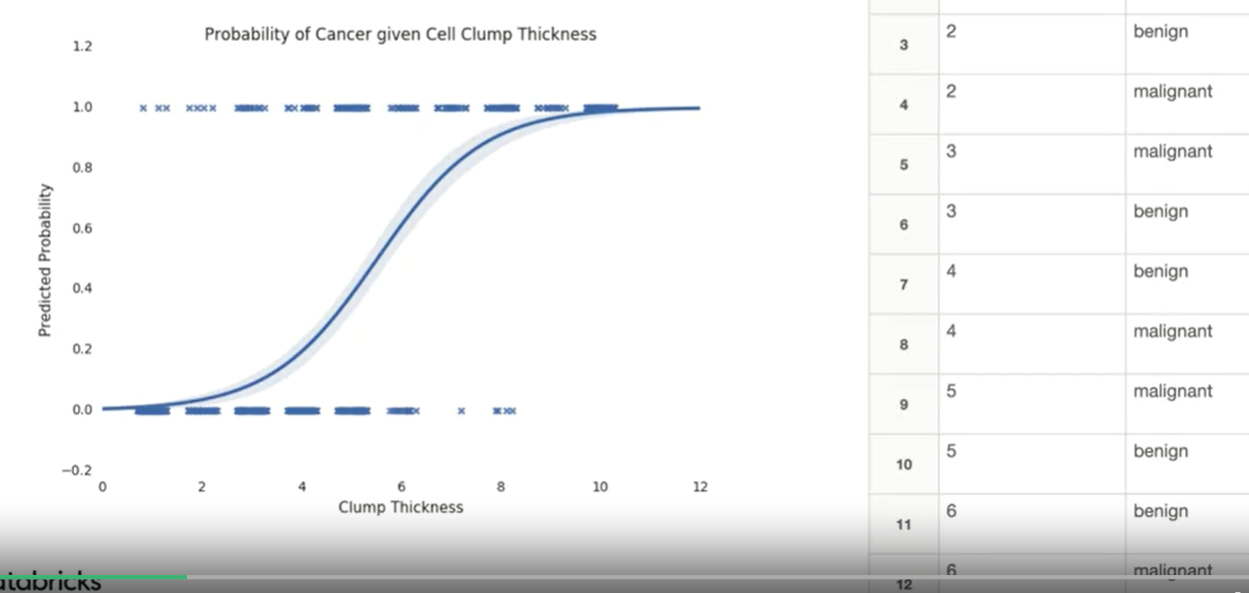
# Applying Logistic Regression



In this video, we'll walk-through an example of how to apply logistic regression to a particular problem. [Bu videoda, lojistik regresyonun belirli bir probleme nasıl uygulanacağına dair bir örnek üzerinden geçeceğiz.]



We'll do this using the Scikit-learn library that we previously learned a little bit about and we'll import the logistic regression model and create a logistic regression object, as you can see in the code block on the slide. [Bunu daha önce biraz öğrendiğimiz Scikit-learn kütüphanesini kullanarak yapacağız ve slayttaki kod bloğunda görebileceğiniz gibi lojistik regresyon modelini import edip bir lojistik regresyon nesnesi oluşturacağız.] Next, we'll identify our input and our output values from our dataset. [Ardından, veri kümemizden girdi ve çıktı değerlerimizi belirleyeceğiz.]



In this dataset our input is the variable clump thickness, which is a numeric variable. [Bu veri setinde girdimiz, sayısal bir değişken olan değişken küme kalınlığıdır.] The output is one of two classes. [Çıktı iki sınıftan biridir.] It has a class label of either benign or malignant. [İyi huylu veya kötü huylu bir sınıf etiketine sahiptir.] These class labels will get coded as zero or one for the sake of our Machine Learning model. [Bu sınıf etiketleri, Machine Learning modelimiz için sıfır veya bir olarak kodlanacaktır.] You can see in the graph on the slide, a reminder of what our logistic function will look like once we train our model on the clump thickness and class data. [Modelimizi küme kalınlığı ve sınıf verileri üzerinde eğittiğimizde lojistik fonksiyonumuzun nasıl görüneceğini hatırlatan slayttaki grafikte görebilirsiniz.] Our model will give us a predicted probability for each sample of whether it's benign or malignant. [Modelimiz, iyi huylu veya kötü huylu olup olmadığına dair her örnek için bize tahmin edilen bir olasılık verecektir.] Next we'll train our model on the training set by fitting it to the x train and y train values and then we can view the coefficients for the model. [Daha sonra modelimizi x dizisi ve y dizisi değerlerine uydurarak eğitim seti üzerinde eğiteceğiz ve ardından modelin katsayılarını örebiliriz.] In this case we have the beta-zero or intercept coefficient. [Bu durumda beta-sıfır veya kesişim katsayısına sahibiz.] Then the beta one, which is the weight that the model gives to the clump thickness coefficient. [Ardından, modelin küme kalınlık katsayısına verdiği ağırlık olan beta bir.] As we can see, it has a pretty high weight ecause in our sample model it is the only variable. [Gördüğümüz gibi oldukça yüksek bir ağırlığa sahip çünkü örnek modelimizde tek değişken o.] That was just a quick walk-through of how you would do a very simple logistic regression model with a dataset that has one input variable and predicts one of two classes for the output variable. [Bu, bir girdi değişkeni olan ve çıktı değişkeni için iki sınıftan birini öngören bir veri kümesiyle çok basit bir lojistik regresyon modelini nasıl yapacağınıza dair hızlı bir gözden geçirmeydi.]