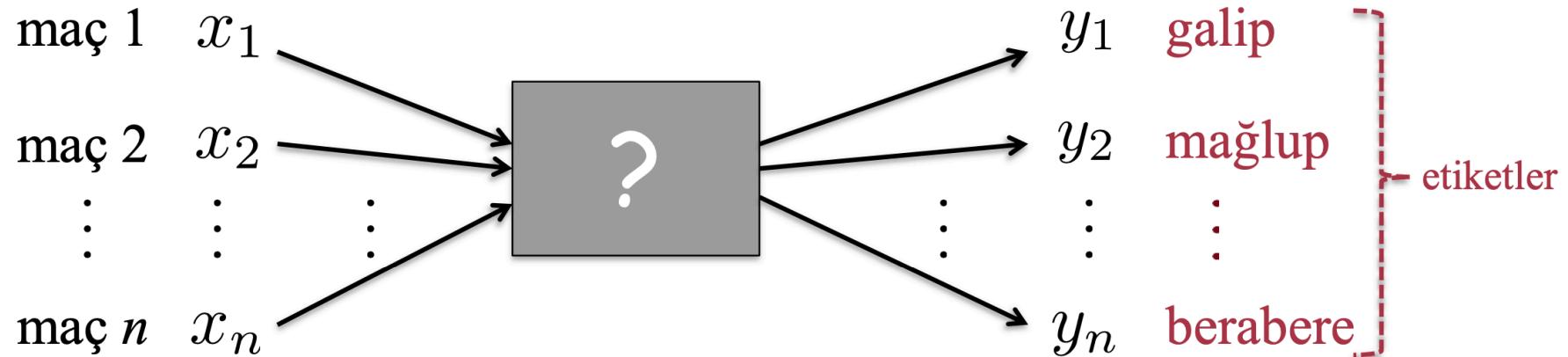




Sınıflandırma - Lojistik Bağlantı

Hesaplamalı Matematik I - Özgür Martin

Sınıflandırma



eğitim verisi
 $\{(x_i, y_i) : 1, \dots, n\}$

Notlar: İlker Birbil, IMO2020, MSGSÜ

Lojistik Bağlanım (Logistic Regression)

X_1 : toplam harcama (bakiye)

X_2 : yıllık gelir

X_3 : öğrenci (1), öğrenci değil (0)

$$Y = \begin{cases} \text{kart borcu kalan (borçlu)}, & 1; \\ \text{kart borcu kalmayan (borçsuz)}, & 0. \end{cases}$$

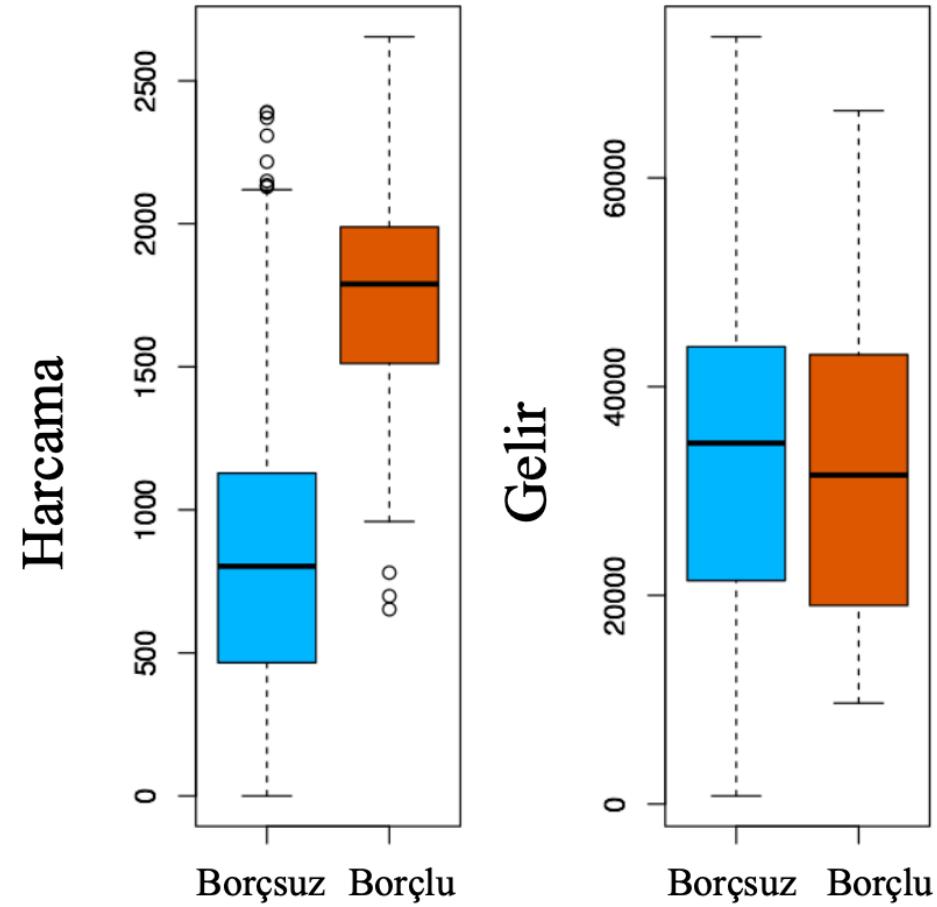
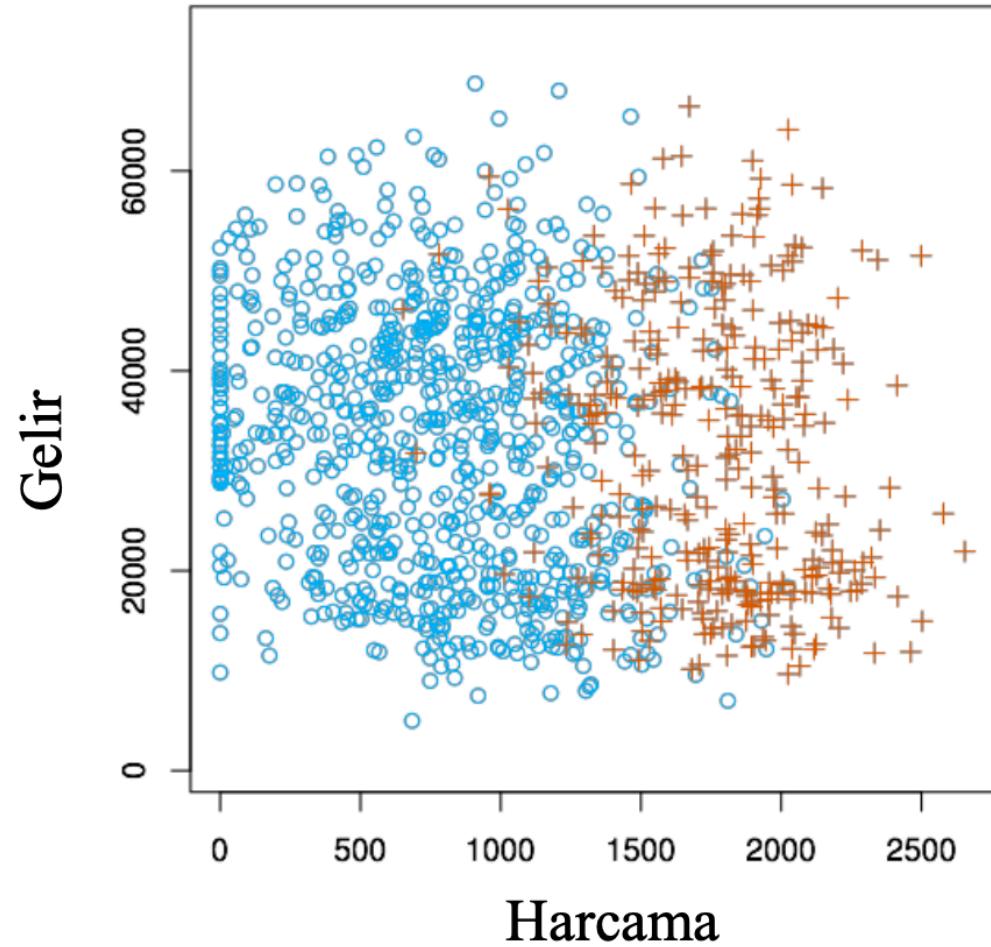
$$p(X) = \mathbb{P}(Y = 1|X) \quad ?$$

$$p(X) = \mathbb{P}(Y = 1|X) > \tau$$



eşik değer (threshold)

Lojistik Bağlanım



$$p(X) = \mathbb{P}(Y = 1|X) \quad ?$$

$$p = 1$$

$$p(X) = \beta_0 + \beta_1 X \rightarrow (-\infty, +\infty)$$

$$p(X) = \sigma(\beta_0 + \beta_1 X) \longrightarrow [0, 1]$$

?

Sigmoid Fonksiyonu

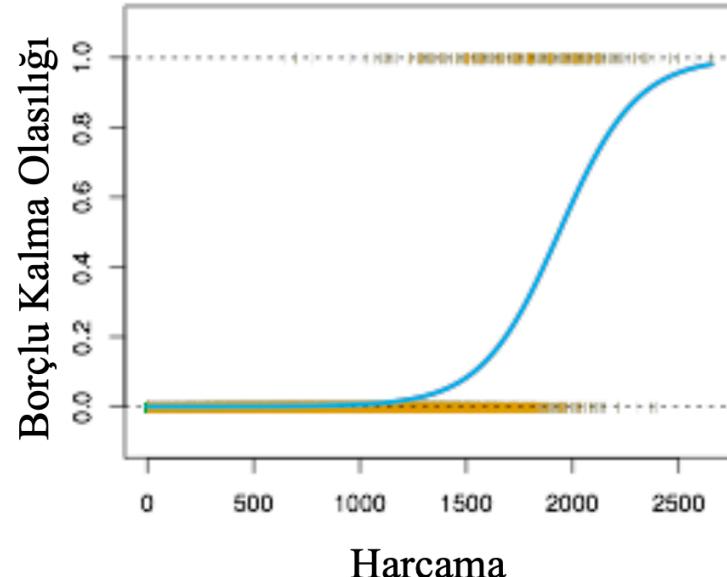
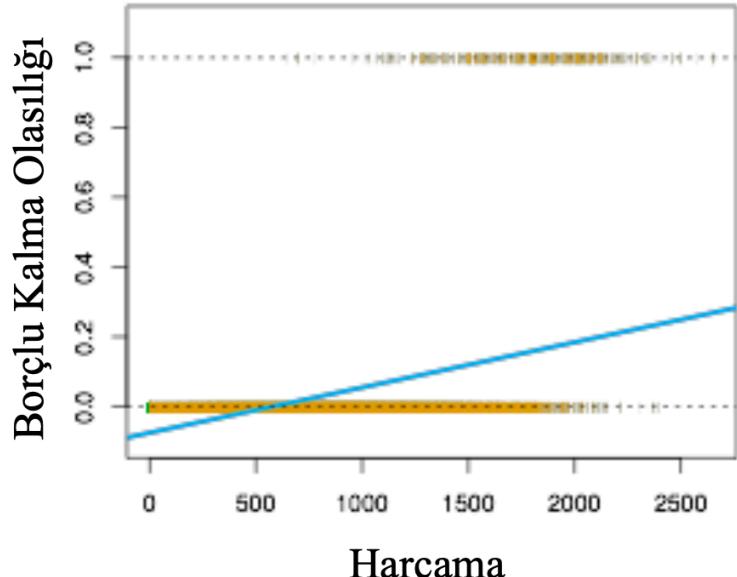
$$\sigma(y) = \frac{e^y}{1 + e^y}$$

Lojistik Fonksiyonu

$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \longrightarrow [0, 1]$$

$$p(X) = \mathbb{P}(Y = 1|X) \quad ?$$

$$p(X) = \beta_0 + \beta_1 X \longrightarrow (-\infty, +\infty) \quad p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \longrightarrow [0, 1]$$



$$\log \left(\frac{p(X)}{1 - p(X)} \right) = \beta_0 + \beta_1 X$$

göreceli olasılıklar (odds)

$$p > 1$$

$$p(X) = \sigma(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)$$

Lojistik Fonksiyonu

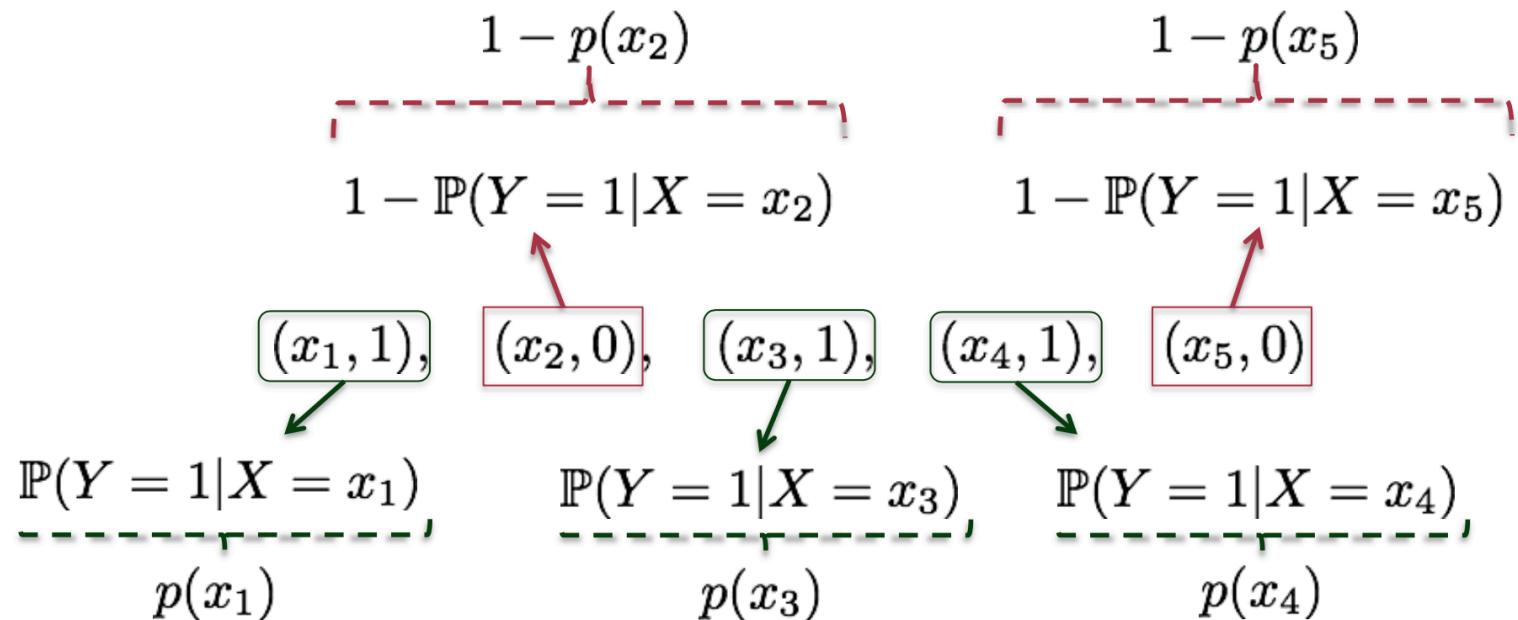
$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}}$$

$$\beta_0, \beta_1, \dots, \beta_p?$$



eğitim verisi

$$\{(x_i, y_i) : 1, \dots, n\}$$



$$p(x_1)p(x_3)p(x_4)(1 - p(x_2))(1 - p(x_5))$$

Olabilirlik (Likelihood) Fonksiyonu

$$\ell(\beta_0, \dots, \beta_p) = \prod_{i:y_i=1} p(x_i) \prod_{j:y_j=0} (1 - p(x_j))$$

En Büyük (Maximum) Olabilirlik

$$\ell(\beta_0, \dots, \beta_p) = \prod_{i:y_i=1} p(x_i) \prod_{j:y_j=0} (1 - p(x_j))$$

$$\max_{\beta_0, \dots, \beta_p} \ell(\beta_0, \dots, \beta_p)$$



$$\max_{\beta_0, \dots, \beta_p} \log(\ell(\beta_0, \dots, \beta_p))$$



$$\max_{\beta_0, \dots, \beta_p} \sum_{i:y_i=1} \log p(x_i) + \sum_{j:y_j=0} \log(1 - p(x_j))$$



İçbükey Fonksiyon



$$\hat{\beta}_0, \dots, \hat{\beta}_p$$

$$p = 1$$

	Katsayı	St. Hata	Z istatistiği	p değeri
Kesme nok.	-3,5041	0,0707	-49,55	< 0,0001
Öğrenci (1)	0,4049	0,1150	3,52	0,0004

$$p > 1$$

	Katsayı	St. Hata	Z istatistiği	p değeri
Kesme nok.	-10,8690	0,4923	-22,08	< 0,0001
Harcama	0,0057	0,0002	24,74	< 0,0001
Gelir	0,0030	0,0082	0,37	0,7115
Öğrenci (1)	-0,6468	0,2362	-2,74	0,0062

Parazit Etkisi (Confounding Effect)

Tahmin Yapma

	Katsayı	St. Hata	Z istatistiği	p değeri
Kesme nok.	-10,8690	0,4923	-22,08	< 0,0001
Harcama	0,0057	0,0002	24,74	< 0,0001
Gelir	0,0030	0,0082	0,37	0,7115
Öğrenci (1)	-0,6468	0,2362	-2,74	0,0062

$$X = \begin{bmatrix} 1.500 \\ 40.000 \\ 1 \end{bmatrix}$$

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

$$\hat{p}(X) = \frac{e^{-10,869 + 0,0057 \times 1.500 + 0,003 \times 40 - 0,6468 \times 1}}{1 + e^{-10,869 + 0,0057 \times 1.500 + 0,003 \times 40 - 0,6468 \times 1}} = 0,058$$

Hata Matrisi (Confusion Matrix)

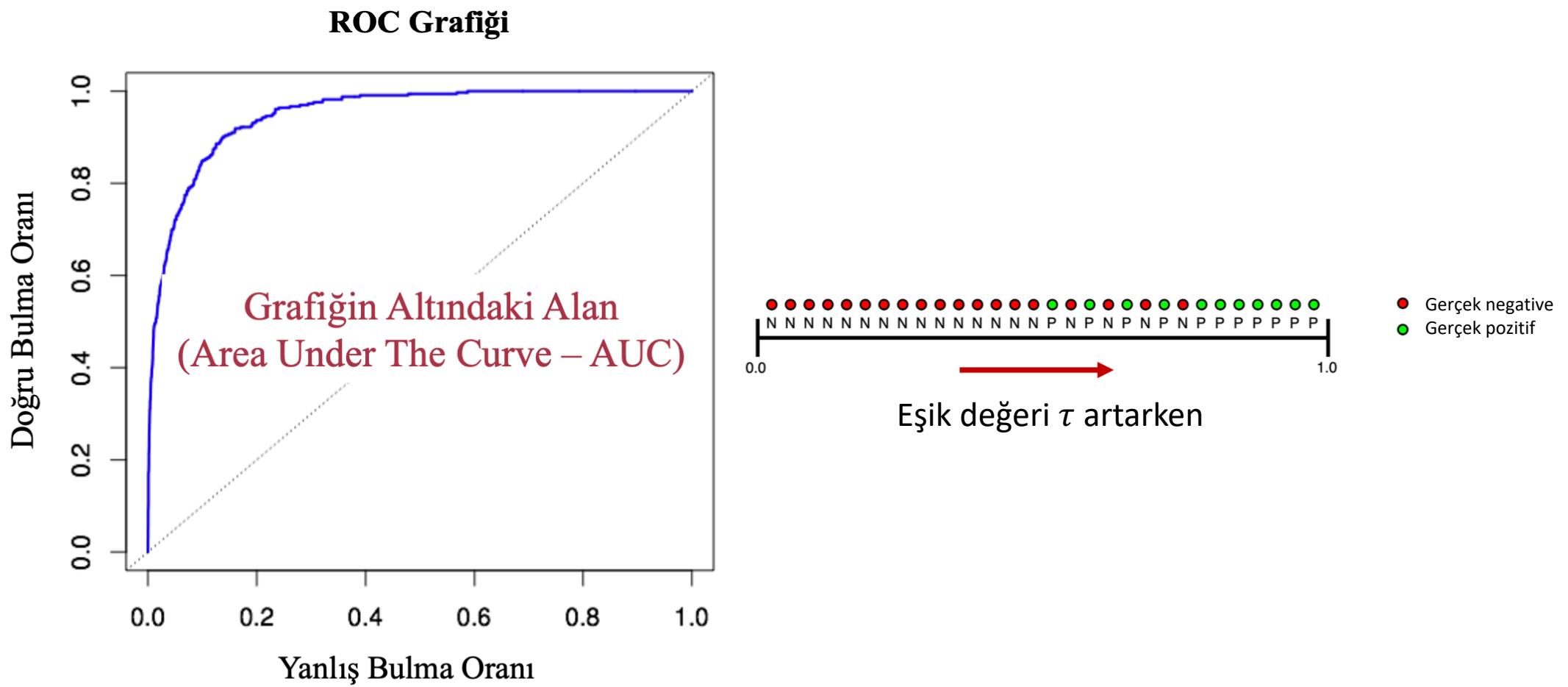
		Tahmin Edilen Sınıf		Toplam
		-	+	
Gerçek Sınıf	-	Gerçek Negatif (GN)	Yanlış Pozitif (YP)	N
	+	Yanlış Negatif (YN)	Gerçek Pozitif (GP)	
	Toplam	N*	P*	

$$p(X) = \mathbb{P}(Y = 1|X) > \tau$$

$\tau = 0.5$		Gerçek		
		Borçsuz	Borçlu	Toplam
Tahmin	Borçsuz	9.644	252	9.896
	Borçlu	23	81	104
	Toplam	9.667	333	10.000

$\tau = 0.2$		Gerçek		
		Borçsuz	Borçlu	Toplam
Tahmin	Borçsuz	9.432	138	9.570
	Borçlu	235	195	430
	Toplam	9.667	333	10.000

ROC* Grafiği (Karar Değerlendirme Grafiği)



* Receiver Operating Characteristics