

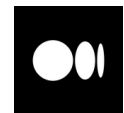
Sınıflandırma Problemlerinde Hata Metrikleri



istdatascience.com | info@istdsa.com



tr.linkedin.com/school/istanbul-data-science-academy



medium.com/istanbuldatascienceacademy

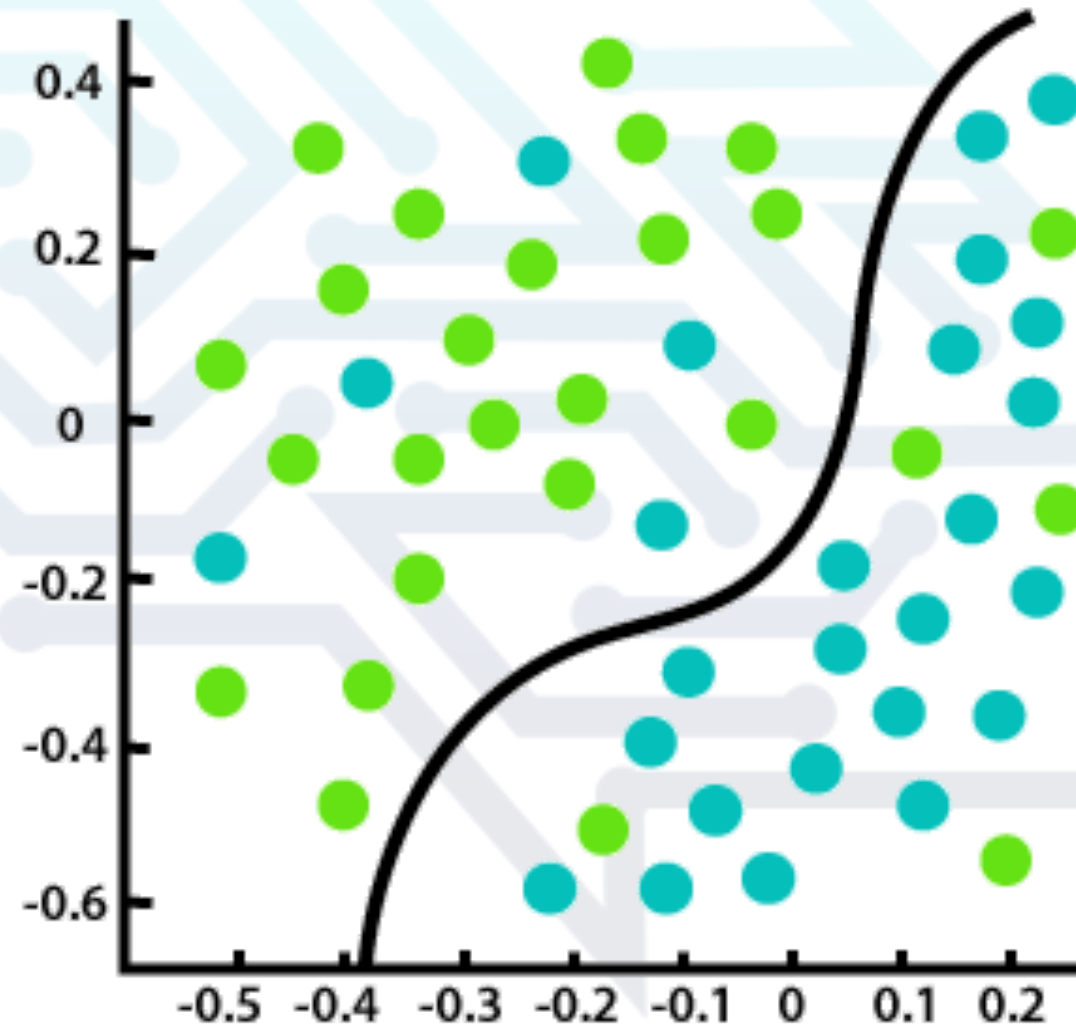
REGRESYON PROBLEMLERİ (HATIRLATMA)

Hata metriklerimizi sınıflandırma problemleri özelinde konuşmaya başlamadan önce, 2. projemizin konusu olan regresyon problemlerinde modellerimizin doğruluğu test etmek için neler kullanıyorduk? Devam etmeden önce en çok kullandıklarımızı hatırlayalım.

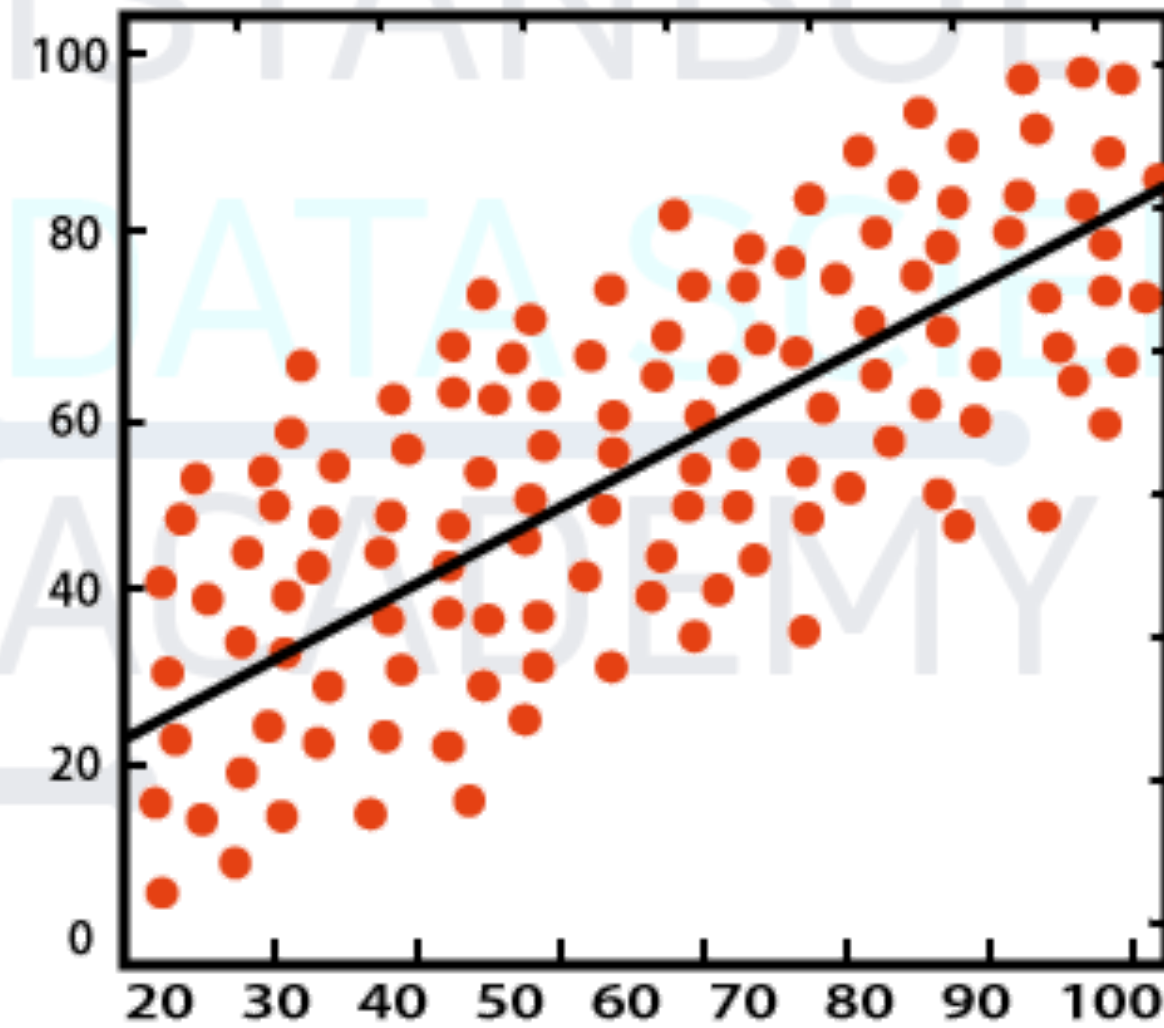
1. **R-squared (R^2)** - Oluşturduğumuz regresyon fonksiyonumuz veri noktalarımızı ne kadar iyi temsil ediyor?
2. **Adjusted R-squared (Adj. R^2)** - Modelimize girdi olarak sağladığımız featurelar gerçekten ne kadar işimize yaramış?
3. **Mean Squared Error (MSE)** - Regresyon fonksiyonumuzla gerçek verilerimiz birbirlerine ne kadar yakın?

SINIFLANDIRMA PROBLEMLERİ

Sınıflandırma problemlerinde ise durum biraz daha farklı, çünkü bu sefer **doğrudan bir sayısal değer tahmini yapmıyoruz**. Elimizdeki verilerin hangi sınıfa ait olduklarını belirlemeye çalışıyoruz. Dolayısıyla daha farklı metriklere ihtiyacımız var.



Classification



Regression

HADI BİR PROBLEMİ ELE ALALIM

Bir ülkede verilen seçim oylarının %54'ünü **demokratlar**, %46'sını **cumhuriyetçiler** almış olsun ve biz, bir seçim çalışması dahilinde seçmen özelliklerine bakarak kimin hangi partiye oy verdiğini tahmin etmek istiyoruz.

- Modelin performansını nasıl ölçebiliriz? Kaç kişiyi **doğru** sınıflandırdığımıza bakarak
- **Accuracy**: Modelimizin tüm sınıfları ne kadar doğru tahmin ettiği bilgisi
- **%95 Model Accuracy**: Bravo! Oldukça iyi bir başarı



BAŞKA BİR PROBLEMİ DAHA ELE ALALIM

Benzer şekilde, bir ülkede yaşayan insanların %99'u sağlıklı, %1'i ise lösemi olsun ve biz, bir tıbbi çalışma dahilinde hasta özelliklerine bakarak kimin lösemi olduğunu tahmin etmek istiyoruz.

- Modelin performansını nasıl ölçebiliriz? Kaç kişiyi **doğru** sınıflandırdığımıza bakarak
- **%99 Model Accuracy:** Tüm insanları sağlıklı şekilde tahmin ediyor
- Bu yeterli bir doğruluk mudur? **Hayır!** Çünkü hiçbir lösemi hastayı bulamadık...
- Demek ki, accuracy bizim tüm ihtiyaçlarımızı karşılamıyor!



CONFUSION MATRIX (HATA MATRİSİ)

		Tahminler	
HATA MATRİSİ		P' (TAHMİN)	N' (TAHMİN)
Olaylar	P (GERÇEK)	TRUE POSITIVE (TP)	FALSE NEGATIVE (FN)
	N (GERÇEK)	FALSE POSITIVE (FP)	TRUE NEGATIVE (TN)

İLK PROBLEME GERİ DÖNELİM

HATA MATRİSİ	DEMOKRAT (TAHMİN)	CUMHURİYETÇİ (TAHMİN)
DEMOKRAT (GERÇEK)	270 (TP)	60 (FN)
CUMHURİYETÇİ (GERÇEK)	110 (FP)	570 (TN)

Accuracy: $(TP+TN) / (TP+TN+FP+FN) = 0.84$

Recall: $TP / (TP+FN) = 0.82$

Specificity: $TN / (TN+FP) = 0.85$

Precision: $TP / (TP+FP) = 0.73$

PRECISION VS RECALL

Precision

Of all **positive predictions**,
how many are **really positive**?

$$\frac{TP}{TP + FP}$$

Recall

Of all **real positive cases**,
how many are **predicted positive**?

$$\frac{TP}{TP + FN}$$

PRECISION: Demokrat dediklerimin gerçekten yüzde kaçını demokrat çıkmış?

$270 / (270 + 110) = \%73$ --> Demokrat tahminlerimin 110 tanesi cumhuriyetçi çıkmış!

RECALL: Demokratların yüzde kaçını doğru olarak sınıflandırmışım?

$270 / (270 + 60) = \%82$ --> 60 tane demokrata cumhuriyetçi demişim!

İKİNCİ PROBLEMİZİ DE DEĞERLENDİRELİM

HATA MATRİSİ	LÖSEMİ (TAHMİN)	SAĞLIKLI (TAHMİN)
LÖSEMİ (GERÇEK)	0 (TP)	10 (FN)
SAĞLIKLI (GERÇEK)	0 (FP)	990 (TN)

Accuracy: $(TP+TN) / (TP+TN+FP+FN) = 0.99$

Recall: $TP / (TP+FN) = 0.00$

Specificity: $TN / (TN+FP) = 1.00$

Precision: $TP / (TP+FP) = \text{TANIMSIZ}$

F1 SCORE: PRECISION & RECALL

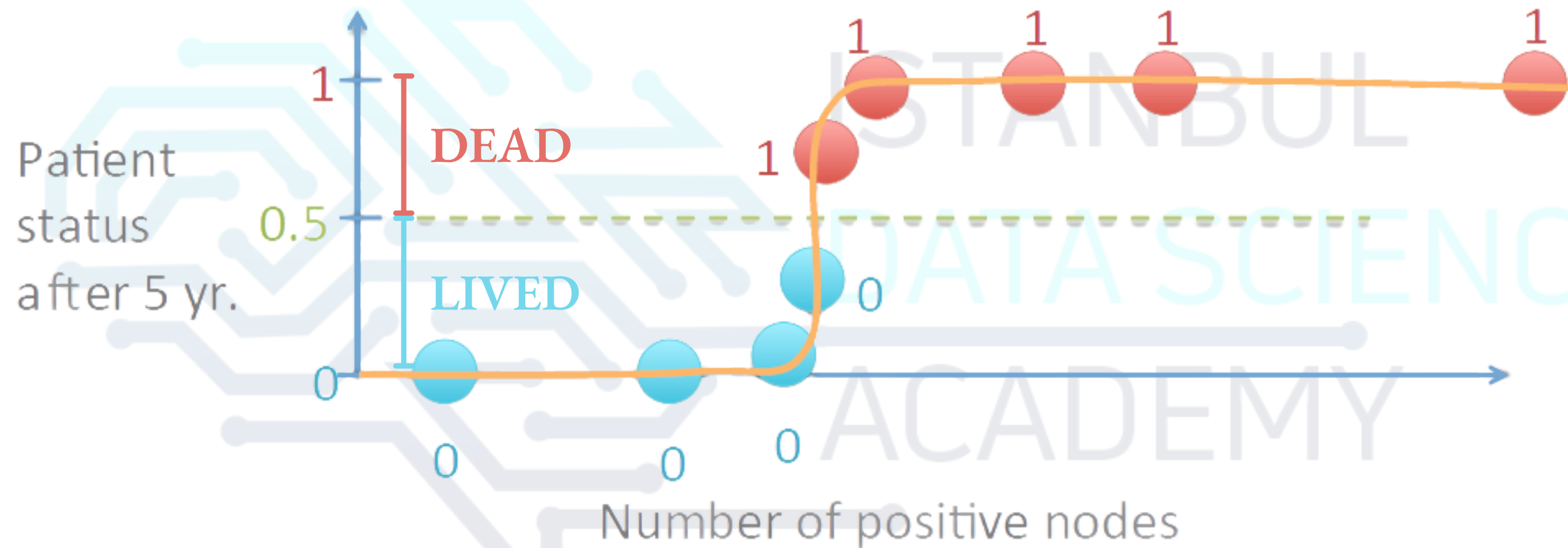
Özellikle tek bir sınıfa dair tahminlerimizi mükemmelleştirmek istediğimiz problemlerde, precision ve recall değerlerini olabildiğince yüksek tutmak isteriz. İşte bu amaçlar için, karşımıza ikisinin harmonik ortalamaları alınarak hesaplanan **F1 Score** değeri çıkar.

Precision: $TP / (TP + FP)$

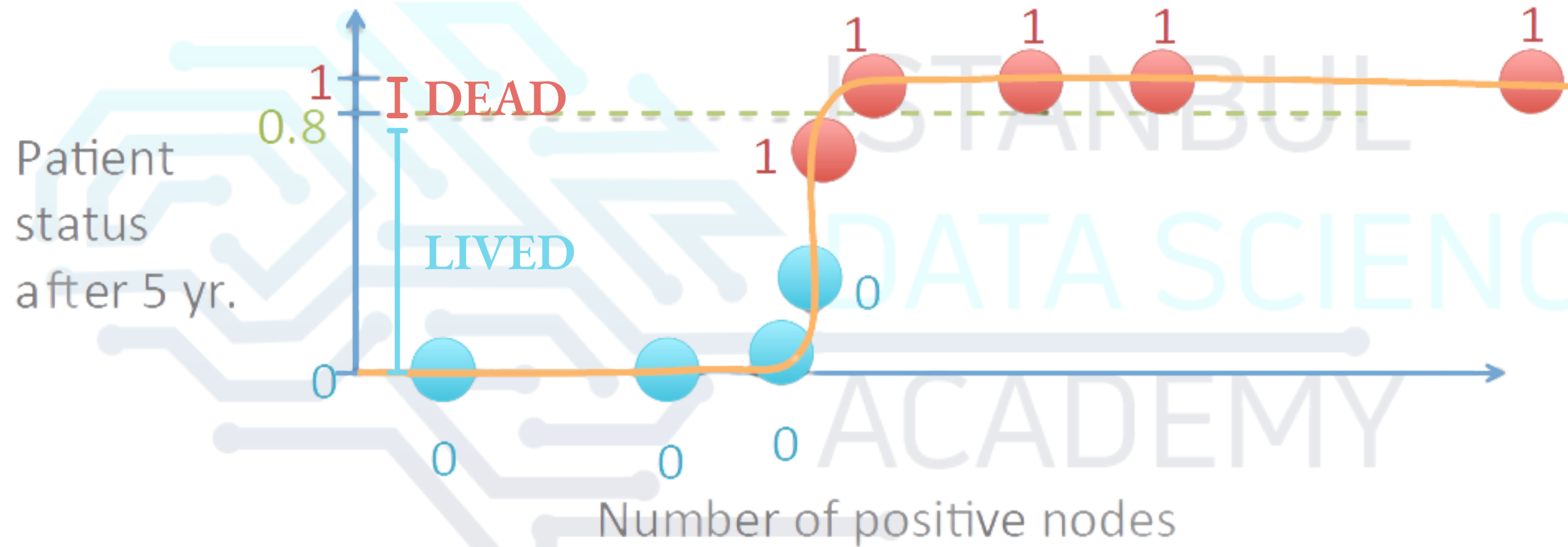
Recall: $TP / (TP + FN)$

F1 Score: $(2 * Precision * Recall) / (Precision + Recall)$

THRESHOLD: OLASILIKLARI YÖNETMEK

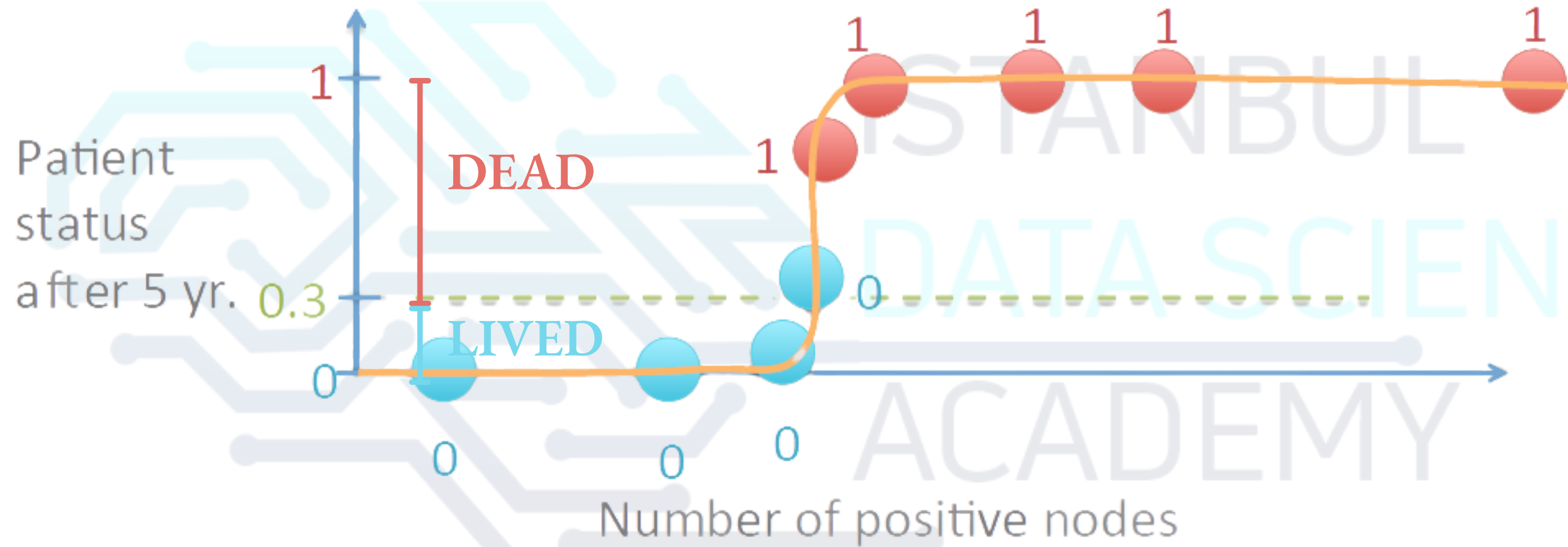


THRESHOLD: OLASILIKLARI YÖNETMEK



- **Yüksek Threshold:** Ölecek olarak yaptığı tahminlerinden daha emin
- **Düşük Recall, Yüksek Precision, Düşük TPR, Düşük FPR**

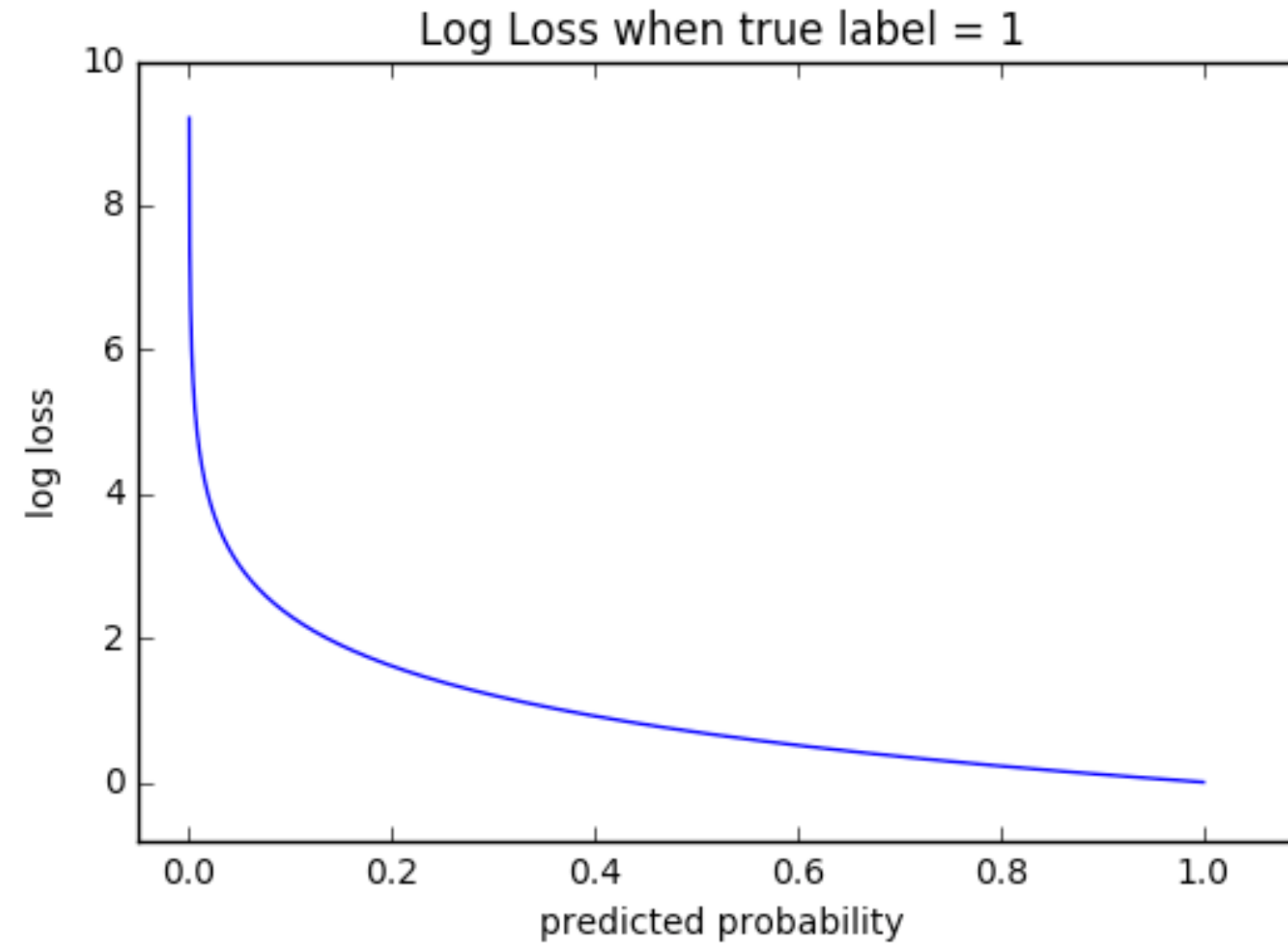
THRESHOLD: OLASILIKLARI YÖNETMEK



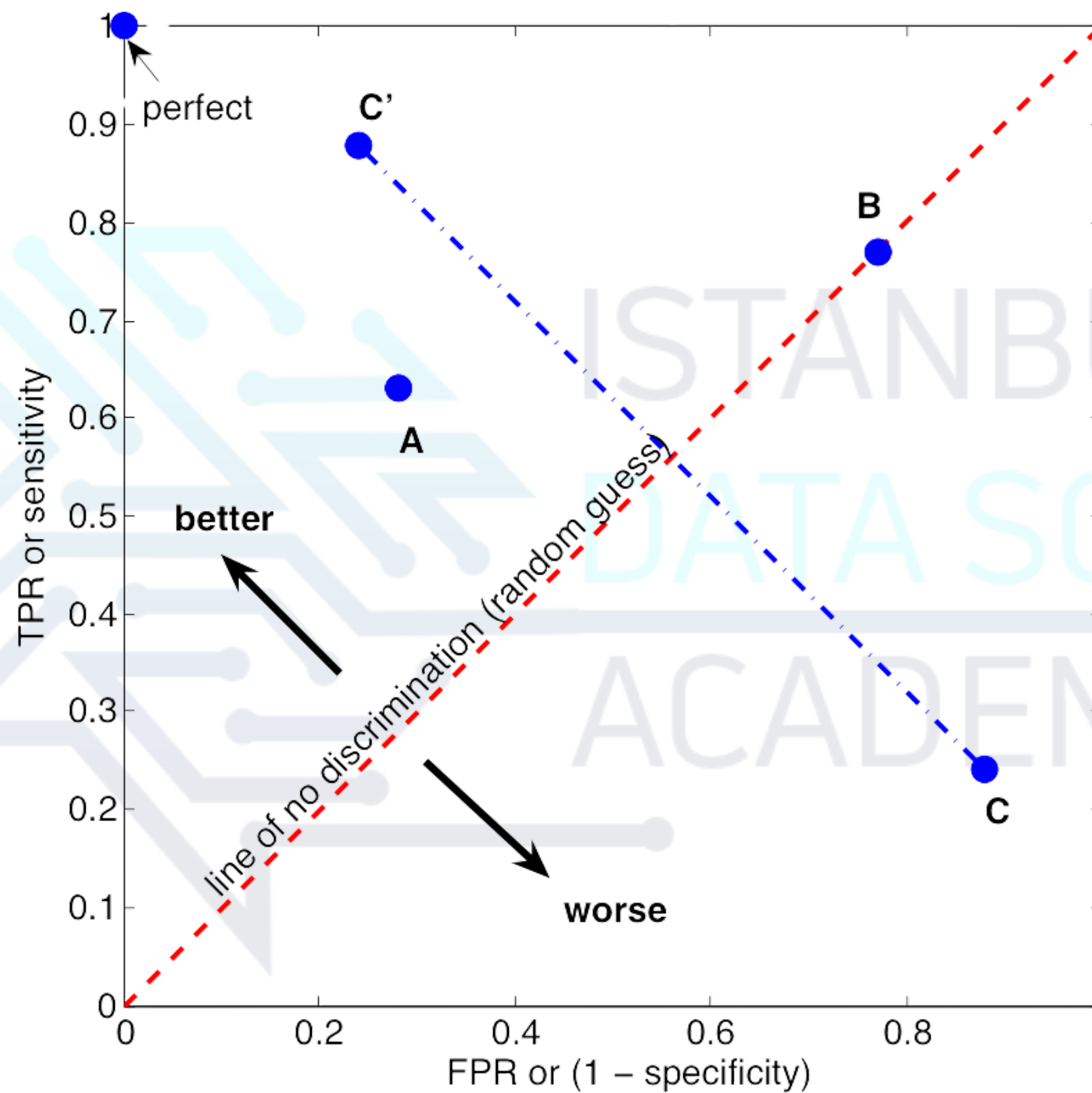
- **Düşük Threshold:** Ölecek olan kişileri yakalamada daha iyi
- **Yüksek Recall, Düşük Precision, Yüksek TRP, Yüksek FPR**

CROSS-ENTROPY

Cross-Entropy veya Log-Loss, çıktısı 0 ile 1 arasında bir olasılık değeri olan bir sınıflandırma modelinin performansını ölçer. Öngörülen olasılık gerçek etiketten ne kadar uzaklaşırsa cross-entropy değeri de o ölçekte artar.

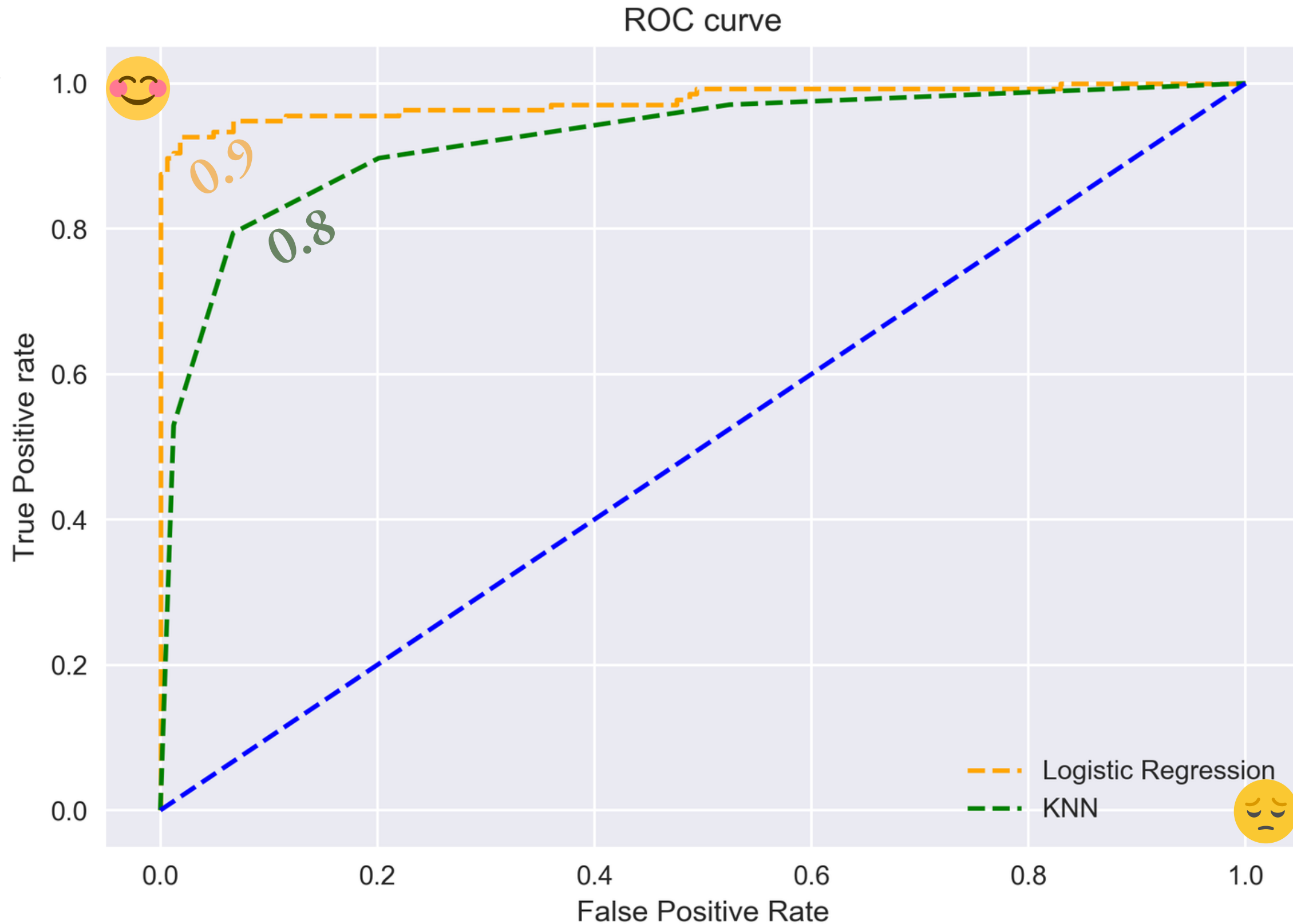


ROC SPACE



AUC (AREA UNDER CURVE)

Yüksek TPR
Düşük FPR



Yüksek TPR
Yüksek FPR

NCE

Düşük TPR
Düşük FPR

Düşük TPR
Yüksek FPR

- Modelimizi eğitim için ayırdığımız training veri setiyle eğitiyoruz.
- Oluşturduğumuz birkaç modeli birbiriyle kıyaslayacaksa, validation veri setiyle amacımız doğrultusunda (**accuracy, precision, recall, F1, AUC** gibi) metriklerden elde edilen sonuçlara bakarak hangi modelin öne çıktığını tespit ediyoruz.
- Elde ettiğimiz sonuçlardan emin olmak için cross-validation çok önemli
- Her şey hazır olduktan sonra, test veri setiyle modelimiz gerçek performansına bakıyoruz.