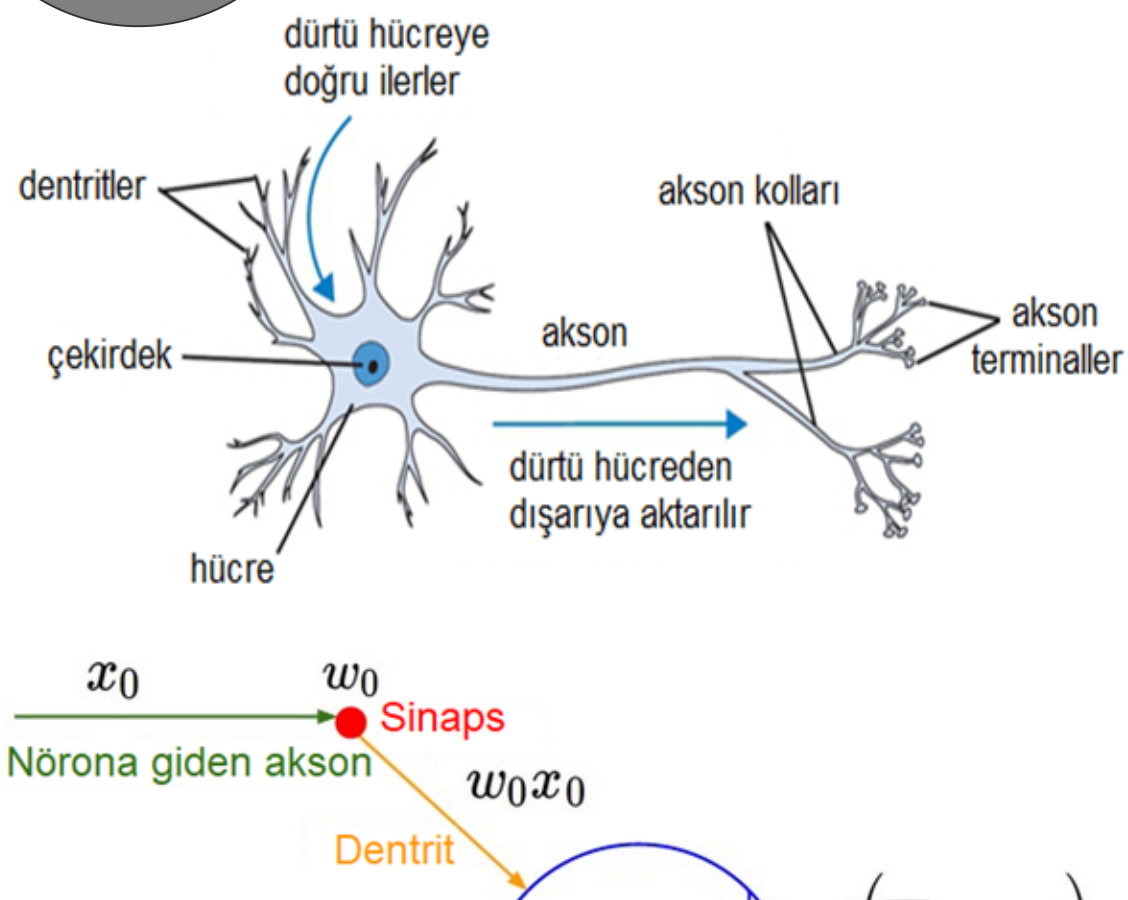


Yapay Öğrenme Model

1

Perseptron (Perceptron): Yapay sinir ağının en küçük parçası olarak bilinen perceptron, aşağıdaki gibi lineer bir fonksiyonla ifade edilmektedir ve ilk defa Frank Rosenblatt tarafından tanımlanmıştır.



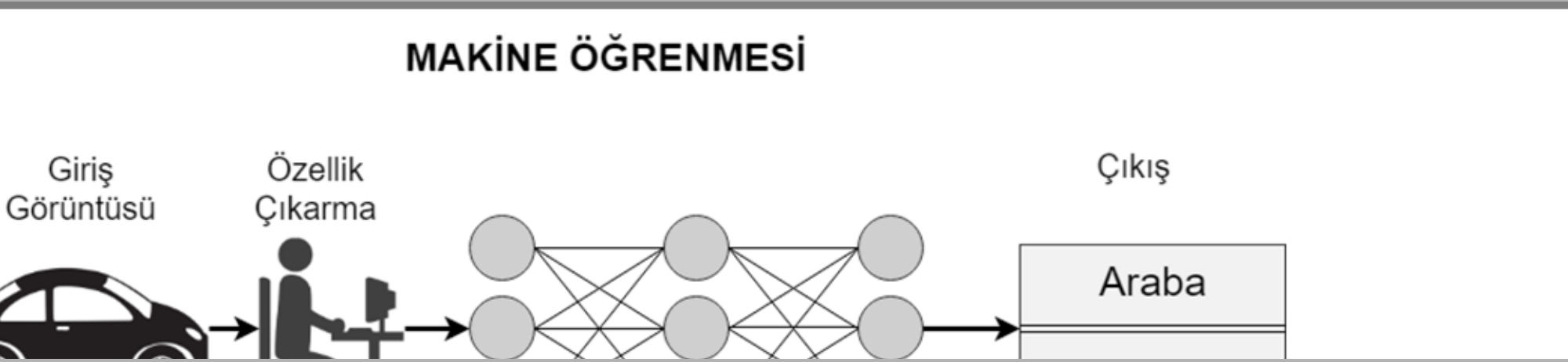
**YAPAY
ÖĞRENME
NEDİR?**

2

Mak
a
s

i Geliştirirken Genellikle Karşılaşıla

ine öğrenmesi dediğimiz 90'lı yıllarda çok yoğun bir şekilde üzerine çalışılan bu
landa "öznitelik çıkarma" (feature extraction) çok önemli bir konuydu. Çünkü
sınırlı sayıdaki veriyi ancak doğru öznitelikler çıkarıldığında ya da seçildiğinde
öğrenebiliyor ve sonuç üretebiliyordu.

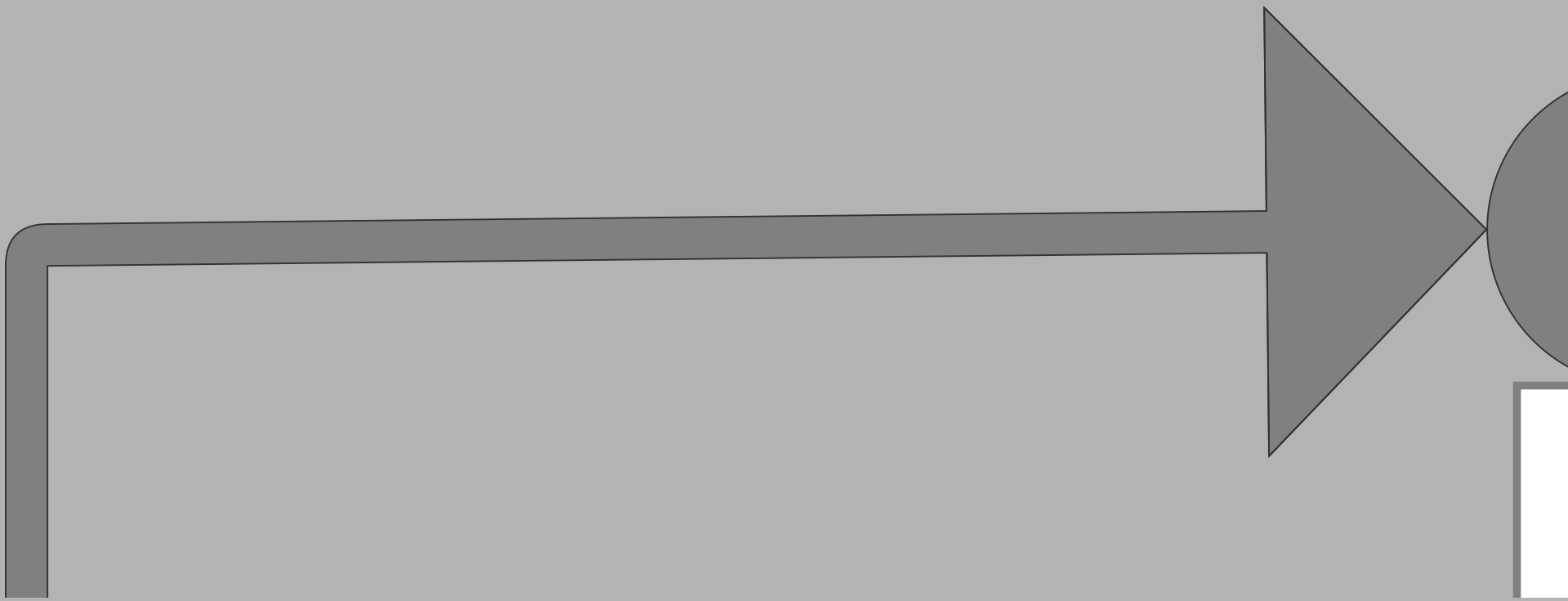


n Problemler ve Olası Çözümleri

İÇİN TEMEL VARSAYIMLAR

nsiyonu için eğitim

nsiyonu için geçerleme



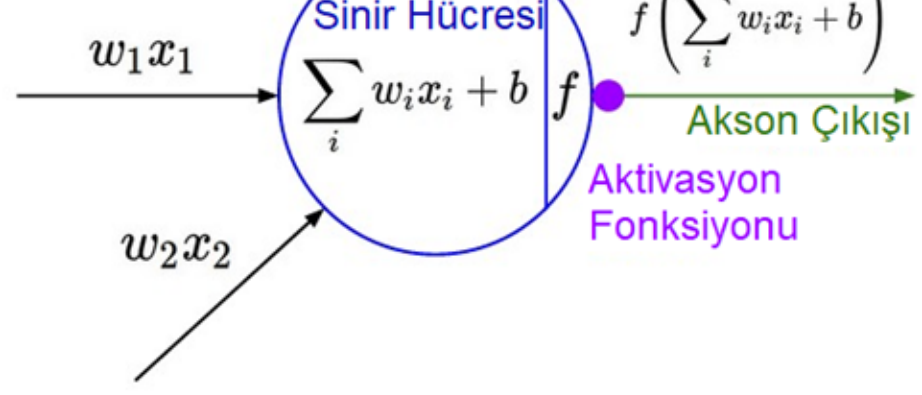
Hazırlayan: Merve Ayyüce KIZRAK
DevNot Summit 2019'da sunuldu

4

**MODELİN MİMARİSİNDEN BAĞIMSIZ BAŞARIMI ARTIRMAK
İÇİN DİKKAT EDİLMESİ GEREKENLER!!!**

Dengesiz veri kümesi problemleri (Problem of Unbalanced Datasets)

Test/Validation/Train Kümesi Verisi



$$y = W \times x + b$$



5

Veri kümesinin yansız olması çok önemli bir konudur. Yani çeşitli sınıflardan olma veride bir ya da birkaç sınıfa yönelik veri çok fazlayken diğerlerinden çok az olma daha çok olduğu sınıfa doğru yanlış bir öğrenme gerçekleştirme ihtimalinizi artırarak istenen bir durum değildir. Örneğin Google çeviri kullandığınızda "hemşire" derken cinsiyetine kadın, "doktor" dediğinde cinsiyetine erkek diye karar vermesinde çok büyük bir önyargı oluşmasına sebep olur. Bu yüzden veri kümelerini büyük ve gerçekçi bir şekilde temsil edecek çeşitlilikte tasarlamak gerekmektedir.



Dengesiz veri kümesi problemleri
(Problem of Unbalanced Datasets)



Less Comprehensive Dataset

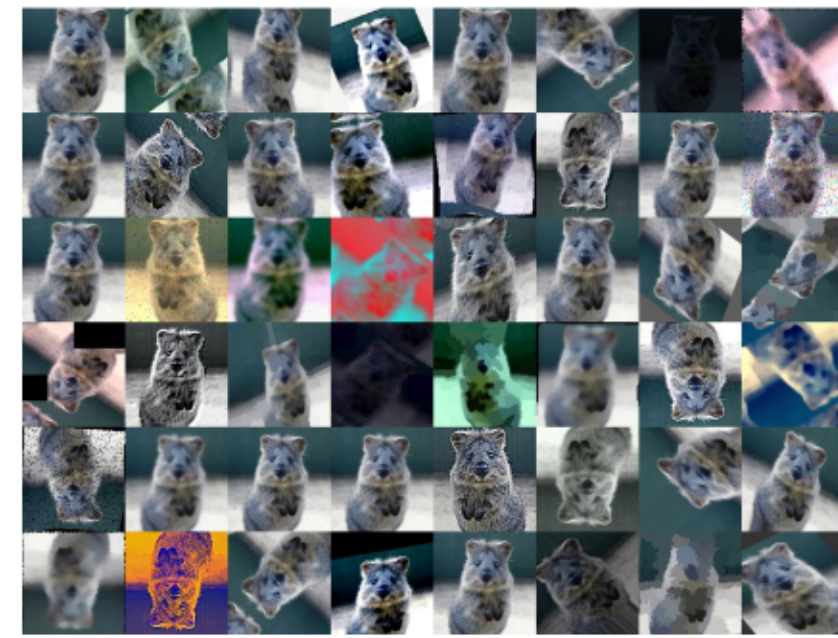


More Comprehensive Dataset



**ÇÖZÜM: VERİ ARTIRMA (DATA AUGMENTATION),
SENTETİK VERİ ÜRETİMİ (SYNTHETIC DATA GENERATION)**

VERİ ARTIRMA
(DATA AUGMENTATION)

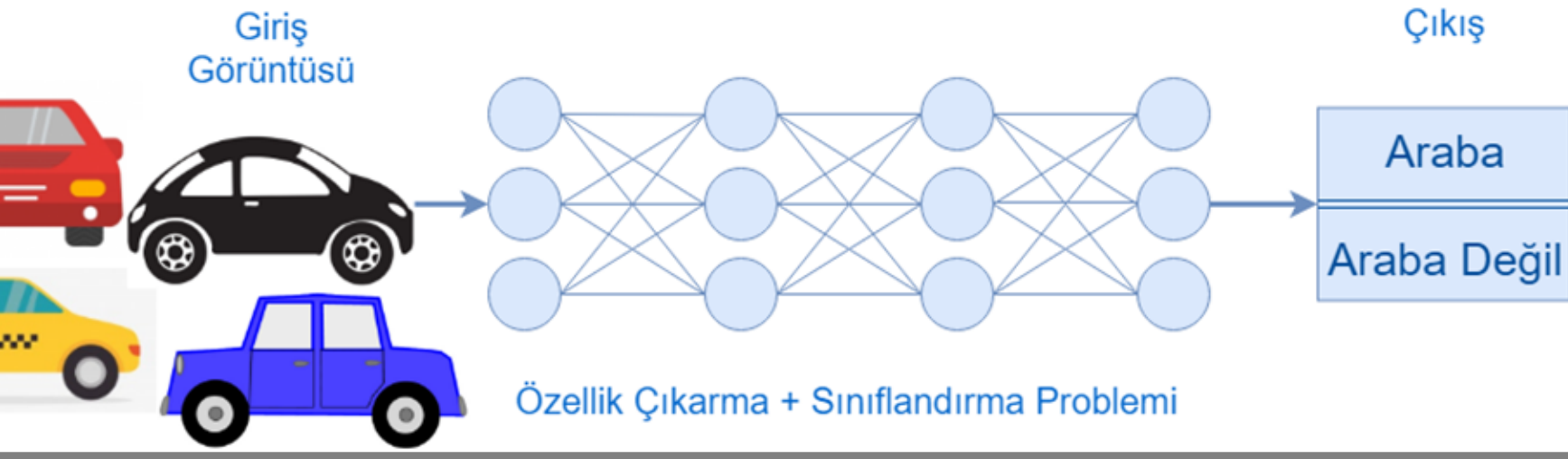


Towards Data Scientist: Image Augmentation for Deep Learning

Sınıflandırma Problemi

Araba Değil

DERİN ÖĞRENME



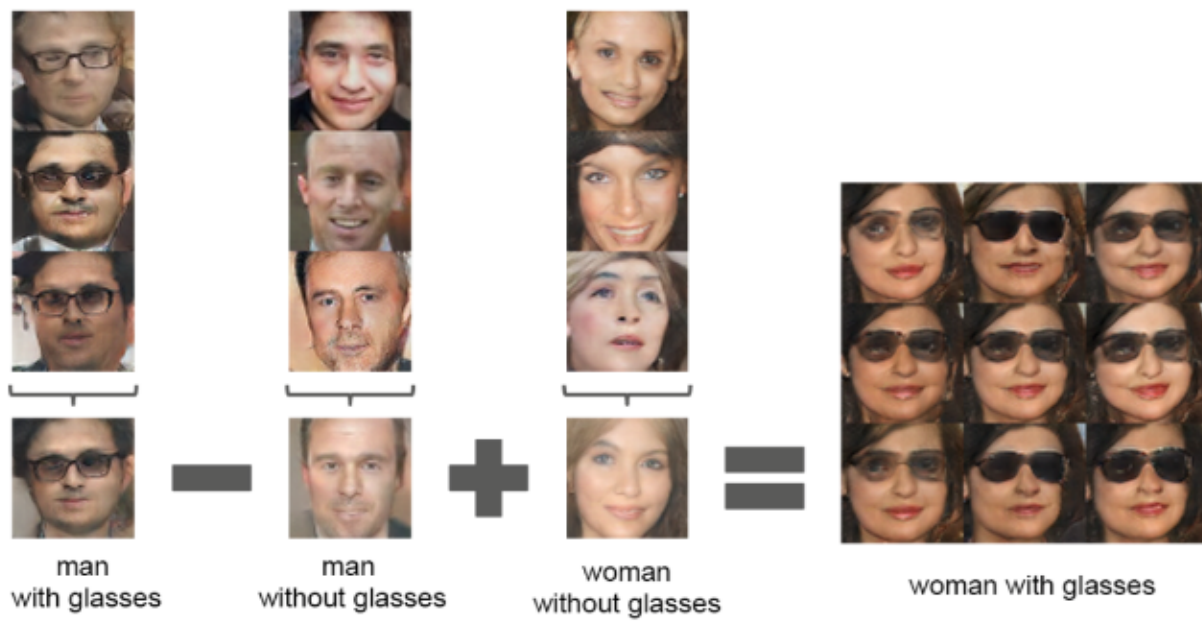
Derin Öğrenme ile aslında temeli 80'li yıllara dayanan ancak 2006 ve 2012 yıllarında Hinton'ın öncülüğünde yayınlanan makaleler ile artık öznelilik çıkarma işlemi de ağıın yetenekleri arasına katıldı. Ama burada önemli olan konu modelin çok daha fazla katmana ve nörona sahip olmasının yanında eğitim verisinin de çok fazla olmasıdır.

Uygun bir maliyet fo

Gerçek dünyada da

uşan bir
ası verinin
rır ve bu
ndiğinde
olduğu gibi
dünyayı

ve SENTETİK VERİ ÜRETİMİ (and SYNTHETIC DATA GENERATION)



arning using Keras and Histogram Equalization

Text

6

Veri kümesi büyüdükçe eğitim için ayrılacak veriyi artırabiliriz. Böylece test kümesinde elde edilen sonuçlar daha iyi öğrenmesine olanak tanıyabilir. Bu durumda %98lere varan eğitim kümeleri mümkündür.

Test/Validation/Train set yapısı

EĞİTİM KÜMESİ

70%

TEST KÜMESİ

EĞİTİM KÜMESİ

60%

GEÇERLEME KÜMESİ

20%

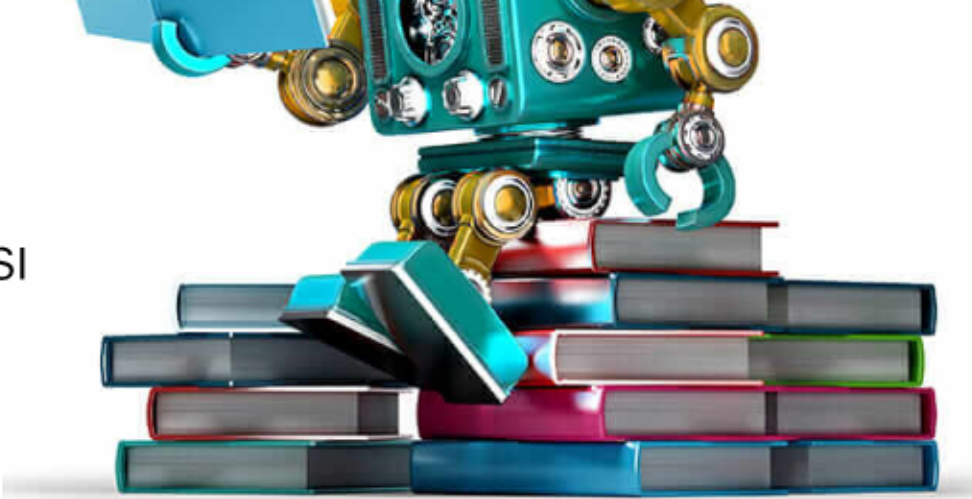
EĞİTİM KÜMESİ

98%

ÇÖZÜM: VERİ KÜMESİNE GÖRE UYGUN

nksiyonu için test

performansının iyi olması



7

Aktivasyon fonksiyonu b
artıran ya da azaltan paan
eğer hem hızlı olsun he
diyorsanız. Gerçekten işi
kadar çok noktada sıfıra g
az bir öğrenme katkı
alındığından (çok katman
daha çok değere sahip

ğımız veri kümesnin % olarak değerini
edilen başarıyı azaltmadan modelin
iz. Veri kümesi milyonlarca veriden
melerinin benimsendiği çalışmalarda

Validation-Geçerleme-Doğrulama kümesi kullanmayı alışkanlık
haline getirmeli, böylece eğitim sırasında öğrenme durumunu
bir yandan kontrol edebiliyor olursunuz!

T KÜMESİ

30%

TEST KÜMESİ

20%

GEÇERLEME

KÜMESİ

TEST KÜMESİ

1

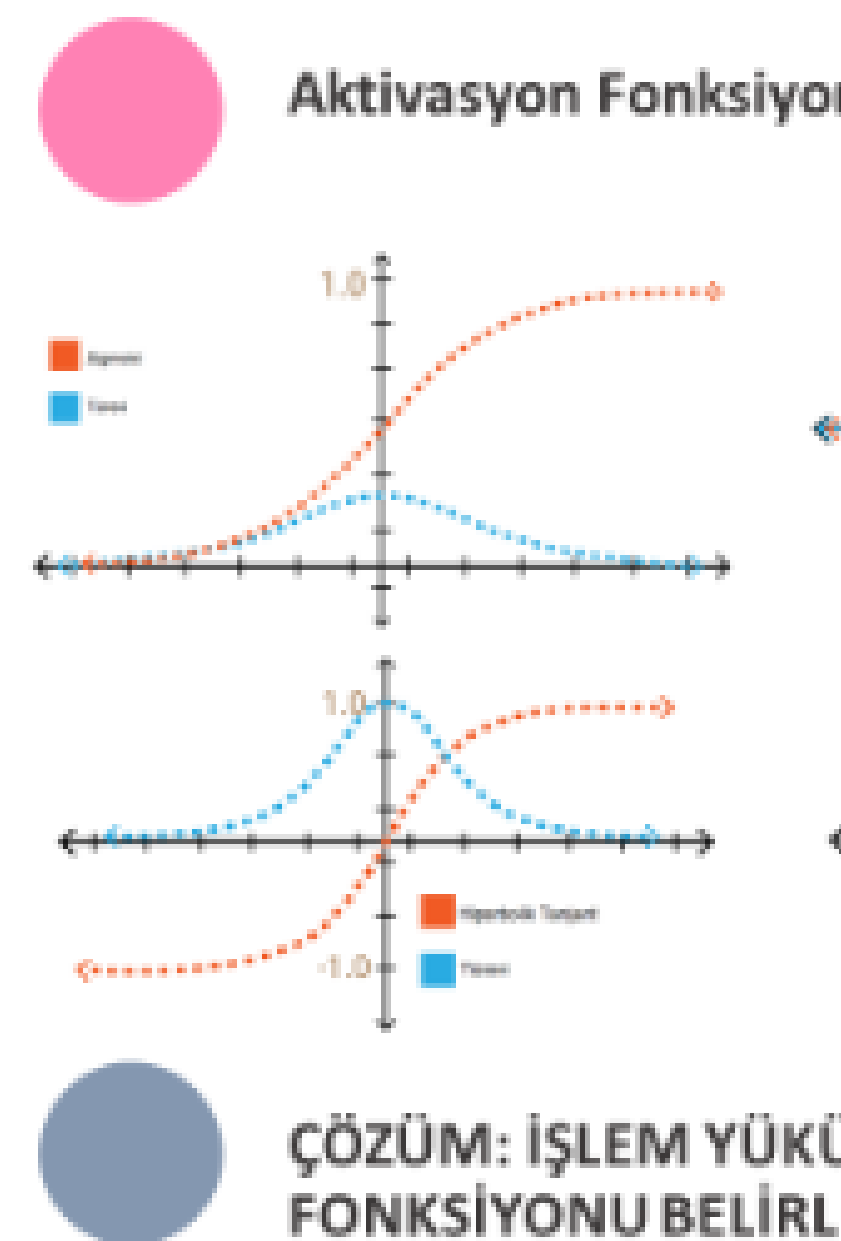
%

1

%

N DAĞILIMLA EĞİTİM VE TEST YAPMAK

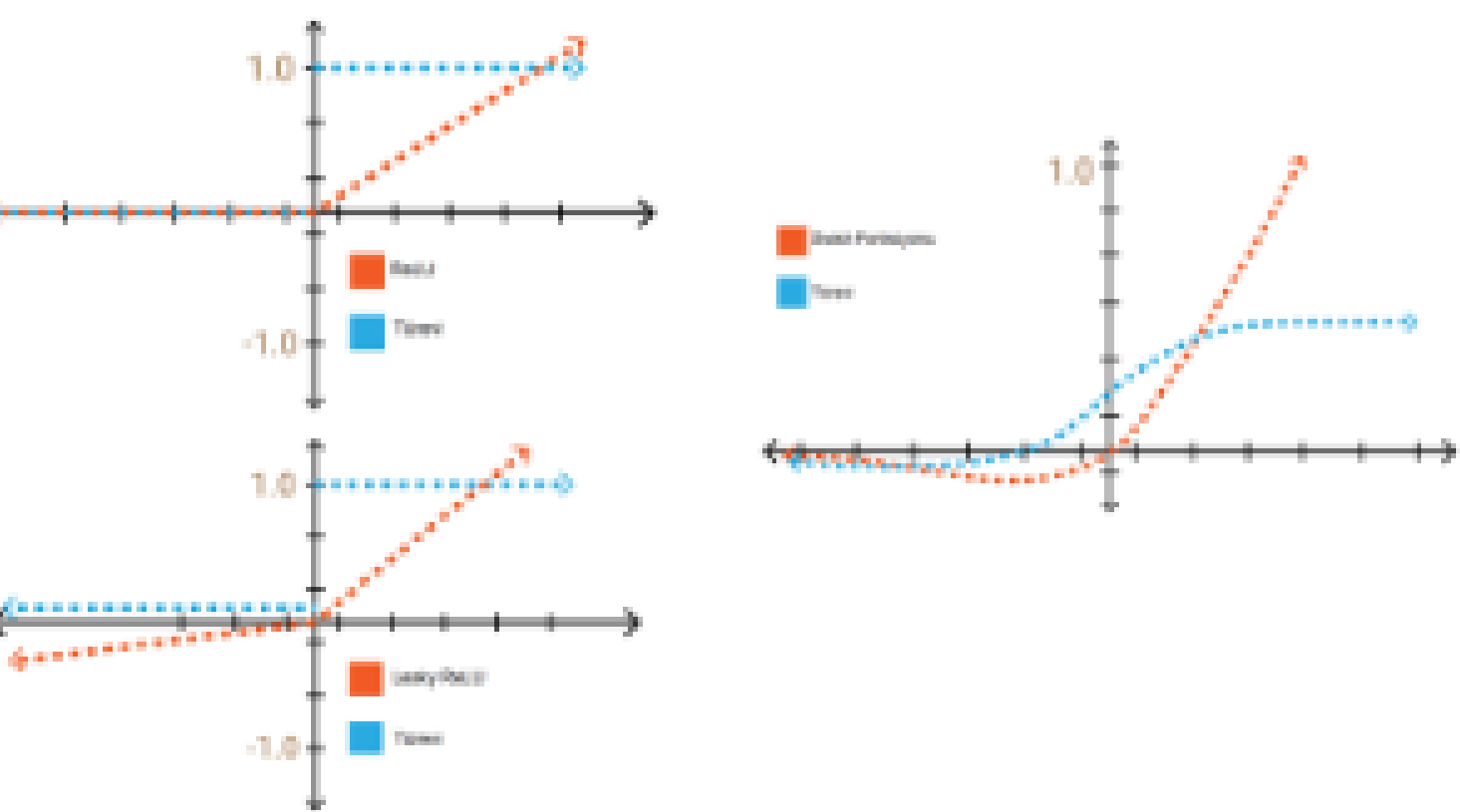
Projects



●	Test/Validation/ Train Kumesi Yapisi
●	Aktivasyon Fonksiyonu Seçimi
●	Optimizasyon Algoritmasi Seçimi
●	Doğruluk Değerlendirme (Accuracy Evaluation)
●	Doğruluk ve Çalışma Süresi İlişkisi
●	Bias vs. Variance
●	Aşırı uydurma- Az uydurma (Overfitting vs. Underfitting)
●	Parametre Hesaplama Yüku

Bir diğ er öğrenme hızını ve oranını etkileyen, işlem yükünü parametrelerden. En çok kullanılanlarına burada yer verildi ancak işlem yükü az olsun hem öğrenmeye katkısı iyi olsun sizin zor. Dikkat etmeniz gereken şey fonksiyonun türevi ne veriyorsa o kadar az işlem yükünüz olur ve o kadar yavaş ve hızlı sağlarsınız. Derin öğrenmede işlem yükü göz önüne alınmadığı için) ReLU daha çok tercih edilir ancak Swish türevde daha iyi olduğu için öğrenmeye katkısı yüksek bir fonksiyondur.

Fonksiyonun Doğru Seçilmesi



YÜKÜ ÇOK ARTIRMADAN EN İYİ ÖĞRENMEYİ SAĞLAYACAK FONKSİYON SEÇMEK

Optimizasyon Yönteminin Seçilmesi

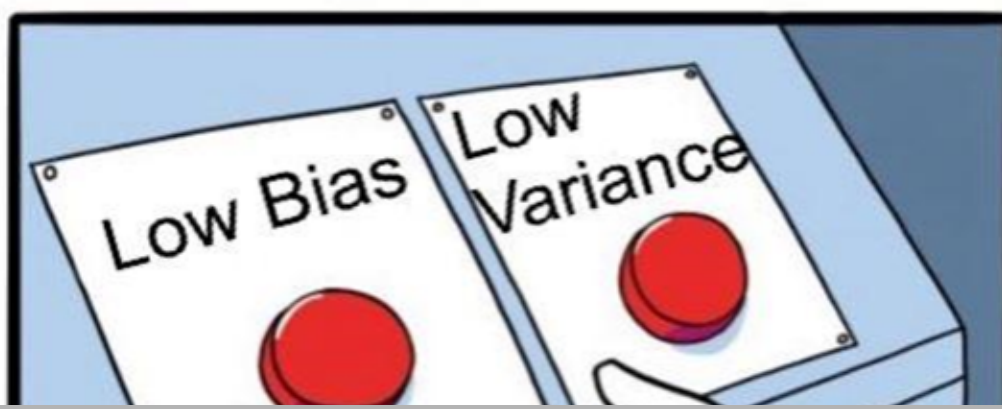


Bir diğer karar verilmesi ve test başarısını doğrudan etkileyen konu da optimizasyon yöntemine karar vermek. Yapay öğrenme modellerinde nihayi hatayı minimize etmek isteriz. Hatayı minimize edecek olan bir optimizasyon algoritması seçememiz gerekir. Elde edilen değerler arasında lokal değil global minimumu ararız. Bazen bu minimuma küçük adımlarla giderken lokal minimumlara takılırız bu istenmeyen bir durumdur. Bir diğer koşulda da büyük adımlar atarız ve global minimumun üstünden atlar geçeriz. Bu da istenmeyen bir durumdur. Optimizasyon yöntemimizi seçmenin kendisinde bir ödünleşim problemi.

ÇÖZÜM: ÖDÜNLEŞİM / LOKAL MİNİMUMA TAKILMADAN HIZLICA GLOBAL MİNİMUMU BULAN YÖNTEM TESPİT EDİLMELİ

11

Bias vs. Variance



Modelin karmaşıklığı - modelin aşırı uydurulması (aşırı öğrenme)

Modelin çok basit olması - modelin az uydurulması (az öğrenme)

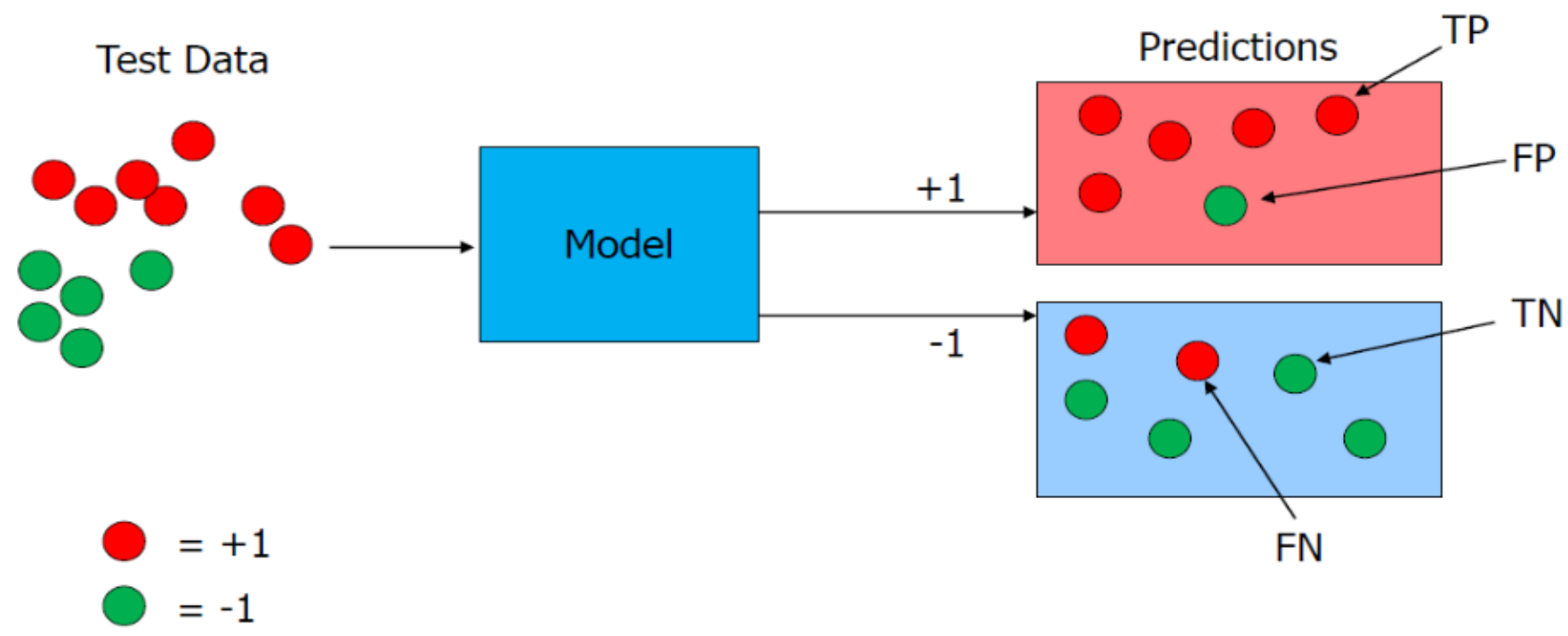
Bu yüzden ortasını bulmalıyız

Bias vs.

Hatası

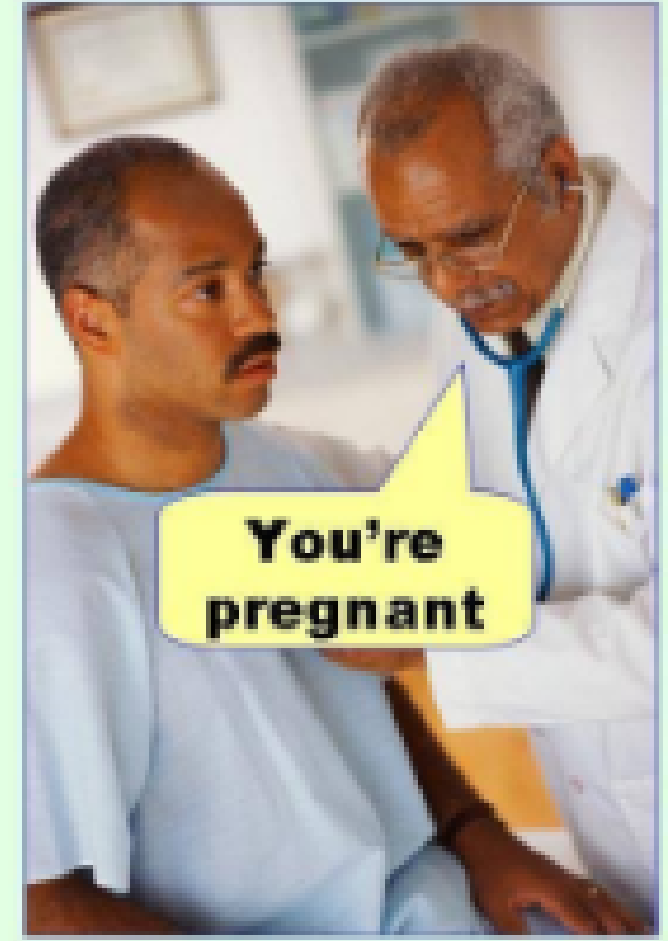
Modelimizin performansını değerlendirikende hangi kriteri seçeceğimize doğru karar vermeliyiz. Doğruluk (Accuracy) başarımla değerlendirme ile bir erkeğe hamilesin diyebilirsiniz ya da yanlış negatif değerlendirmesi ile hamile olduğu aşikar bir kadına hamilelik metriklerini de incelemek modelinizin gerçek başarımlarını

Doğruluk Değerlendirme (Accuracy Evaluation)



ÇÖZÜM: FARKLI PERFORMANS KRİTERLERİNE GÖRE DE DEĞERLENDİRME YAPMAK

Type I error
(false positive)



Type II error
(false negative)



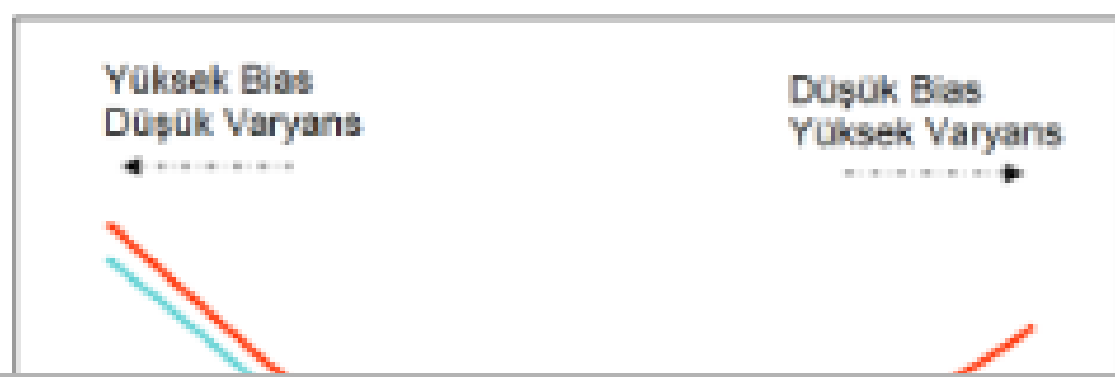
BIAS - VARIANCE

Öğrenmesi ya da overfitting dediğimiz durum) ve açıklanabilirliğini kaybetmesiyle sonuçlanmayan durumlardandır.

Öğrenmesi ya da underfitting dediğimiz durum) matematiksel olarak kolayca öğrenilebilir ve genelleştirilebilir. Fakat başarısı da bir o kadar düşüktür.

! Yolu takip edin nasıl olduğunu göreceksiniz...

Variance



NEDEN BASİT MODEL SEÇMELİYİZ?

- Kullanımı basittir – Hesaplama karmaşıklığı düşüktür!
- Eğitilmesi kolaydır – Alan gerekliliği düşüktür!



m kriterlerinden yalnızca biridir. Öyle durumlar vardır ki doğruluk başarımınız yüksekken doğru yanlış pozitif bir hamile değilsin teşhisi verebilirsiniz. :) Buyüzden hassaslık (sensivity) recall (hatırlama) gibi farklı performans arısını ölçmenizde çok fayda sağlar.



Doğruluk Değerlendirme (Accuracy Evaluation)

Karşılık Matrisi/ Olasılık Tablosu	Gerçek Sınıflar (Gözlem)	
	tp (true positive) Gerçek Pozitif	fp (false positive) Beklenmedik Sonuç
Kestirim Sınıfı (Beklenen)	fn (false negative) Kaybedilen Sonuç	tn (true negative) Gerçek Negatif

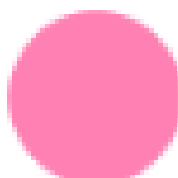
Precision = $\frac{tp}{tp+fp}$
(Hasasiyet)

Accuracy = $\frac{tp+tn}{tp+tn+fp+fn}$
(Doğruluk)

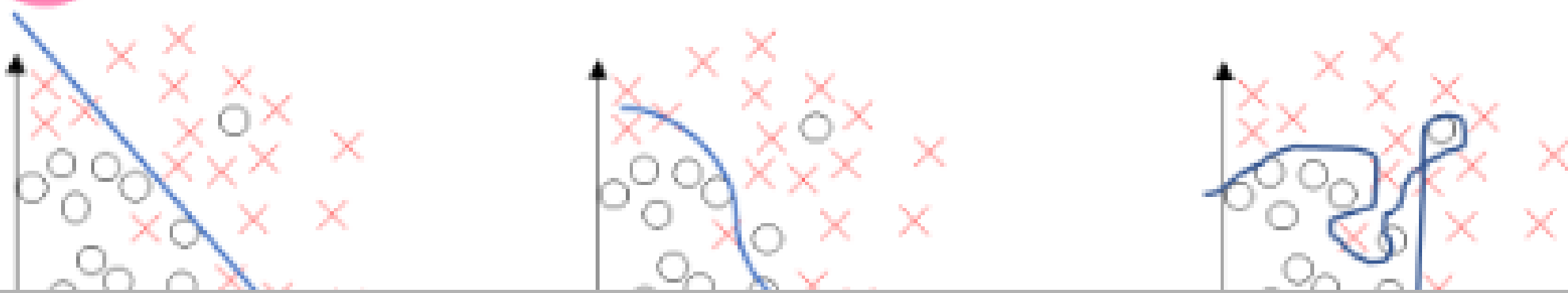
Recall = $\frac{tp}{tp+fn}$
(Kesinlik)



ÇÖZÜM: FARKLI PERFORMANS KRİTERLERİNE GÖRE DE DEĞERLENDİRME YAPMAK



Bias vs. Variance



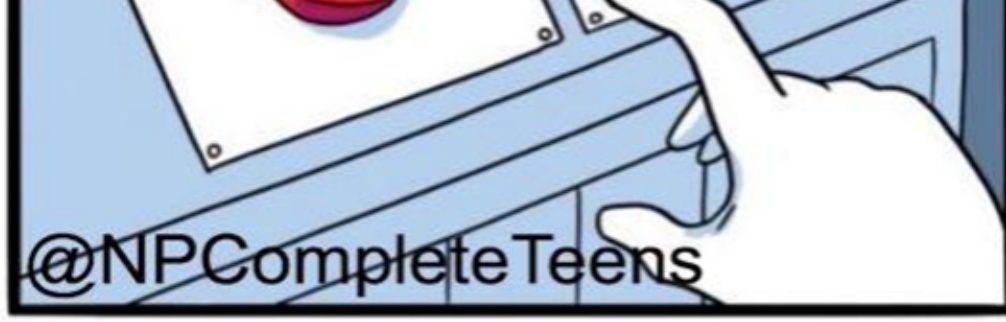
Doğruluk ve Çalışma Süresi İlişkisi

Classifier	Accuracy	Running time
A	90%	80ms
B	92%	95ms
C	95%	1,500ms

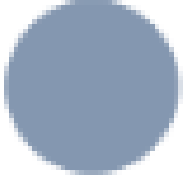
ÇÖZÜM: ZAMANIN ÖNEMLİ OLDUĞU DURUMLARDA BAŞARIMDAN ÖDÜN VERİLEBİLİYORSA BU YOL SEÇİLEBİLİR.

<https://www.deeplearning.ai/Course 3 Structured Machine Learning Projects>

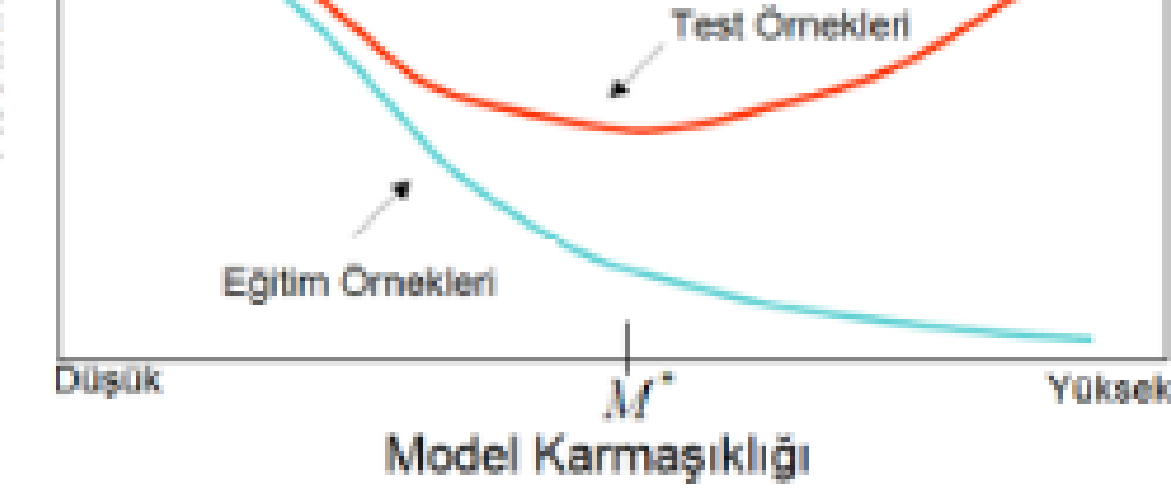
Bazen de hız önemli olur! Bir model daha iyi öğrenip daha yüksek başarı gösterirken yavaş çalışır. Bir diğer model ise çok hızlıdır ancak o kadar başarılı değildir. Bu da yine sizin neye ihtiyacınız olduğuna göre karar vermeniz gereken bir konudur. Başarının biraz düşmesi sizi çok etkilemiyorsa hızınızı artırmayı tercih edebilirsiniz.



ÇÖZÜM: GENELLEŞTİRİLEBİLİR, AÇIKLANABİLİR VE BAŞARILI BİR MODEL

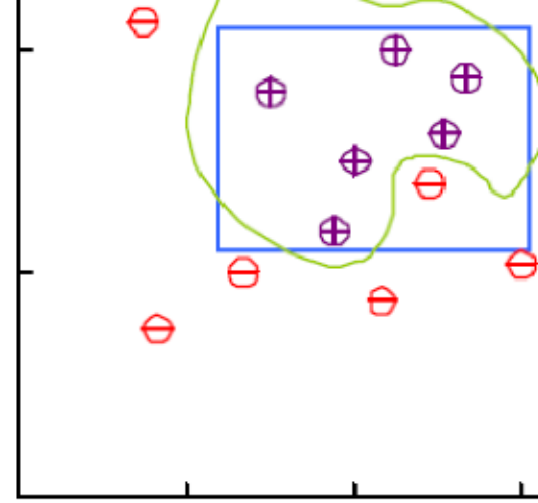


ÇÖZÜM



İ: GENELLEŞTİRİLEBİLİR, AÇIKLANABİLİR VE BAŞARILI BİR MODEL

- Açıklaması kolaydır – Yorumlanabilir!
- Genelleştirilebilirdir!



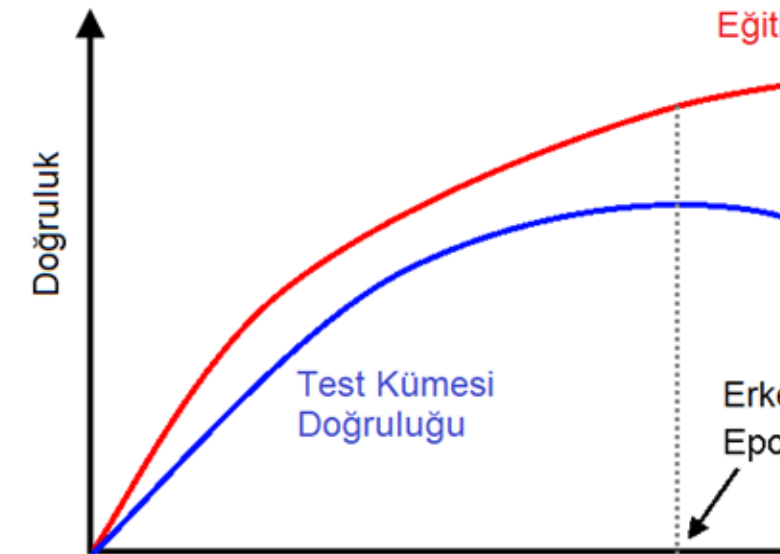
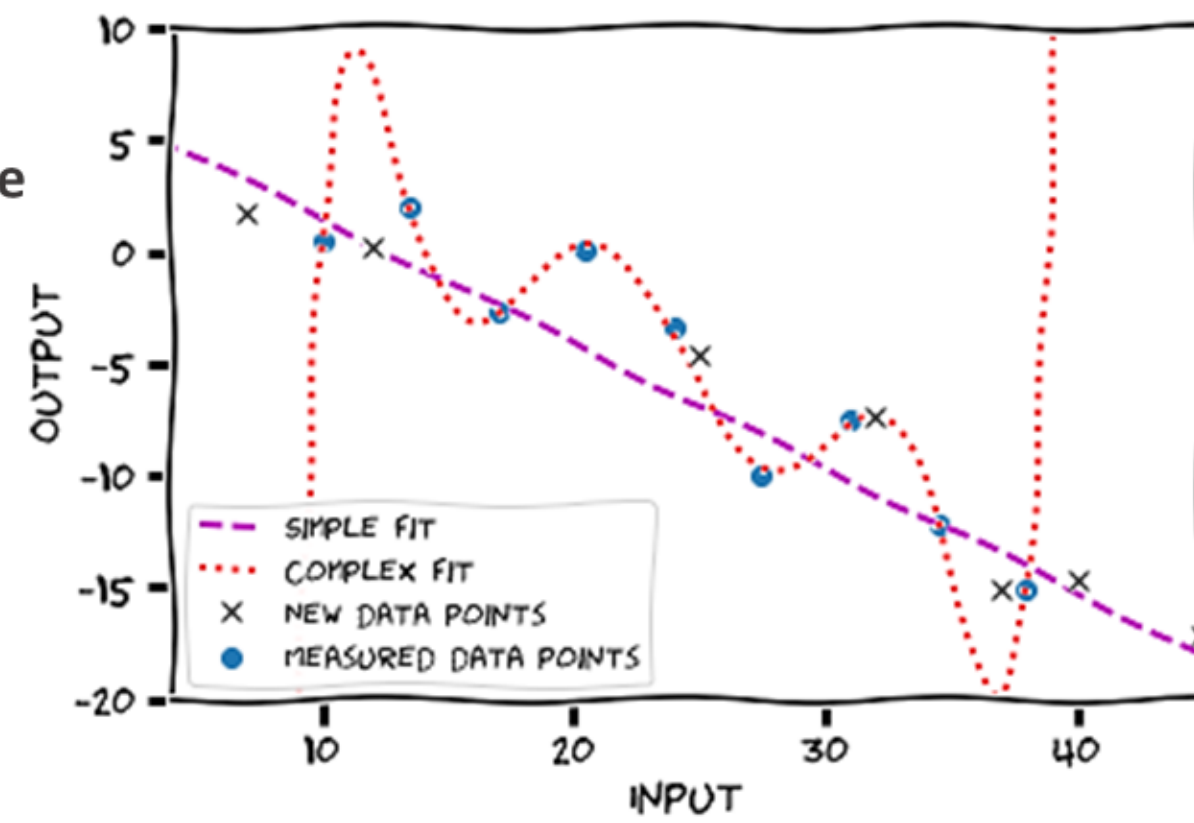
Yakup Genç, GTU Derin Öğrenme Ders Notları

OVERFITTING NASIL ÖNLENEBİLİR SORUSUNUN YANITI:

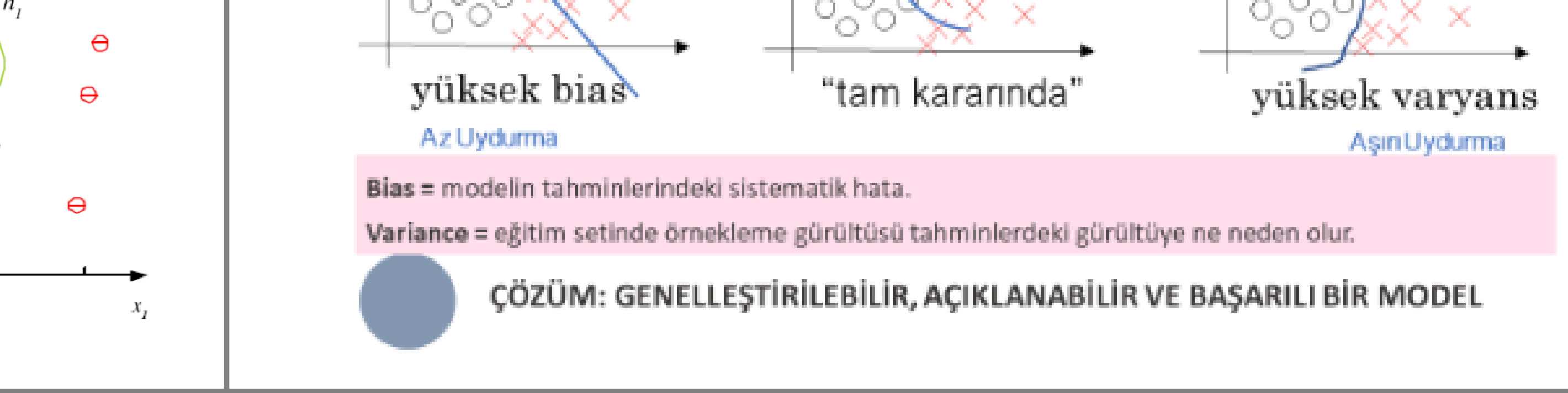
- VERİYE GÜRÜLTÜ EKLEMELER OLABİLİR GERÇEK DÜNYAYI TEMSİL EDEBİLMESİ VE GENELLEŞTİRİLEBİLİR HALE GETİRİLMESİ
- DROPOUT DEDİĞİMİZ SEYRELTME İŞLEMİ İLE BAZI NÖRONLARIN SİLİNMESİ İLE AŞIRI ÖĞRENMEYİ AZALTIP EĞİTİMİ DÜŞÜREBİLİRİZ.
- BİR DİĞER YÖNTEM DE "ERKEN DURDURMA" BÖYLECE TEST BAŞARISI DÜŞMEYE BAŞLADIĞINDA EĞİTİM DURDURULUP YENİ VERİLERLE TEST YAPILABİLİR. BU UYGULANABİLECEK EN KOLAY YÖNTEMLERDENDİR.

Gürültü Ekleme İşlemi

- Modelin girdilerine gürültü ekleme
- Modelin gizli katmanlarına gürültü ekleme
- Modelin ağırlık parametrelerine gürültü ekleme:
- Loss fonksiyonuna, gürültüyü parametre olarak ekleme



ÇÖZÜM: SEYRELTME



İLMESİ İÇİN

M HIZINI DA AYNI ZAMANDA ARTIRABİLİRİZ.

LUR BÖYLECE EN YÜKSEK BAŞARIDA TEST İŞLEMİ

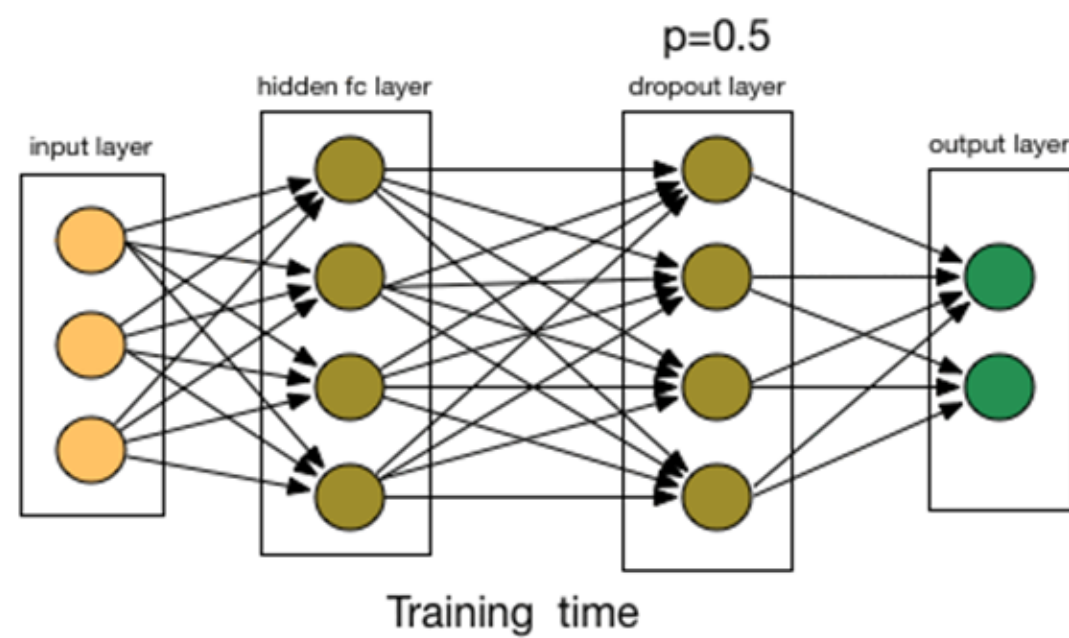
a- Az uydurma (Overfitting vs. Underfitting)

im Kümesi Doğruluğu

Aşırı Uydurma

en Durdurma
och

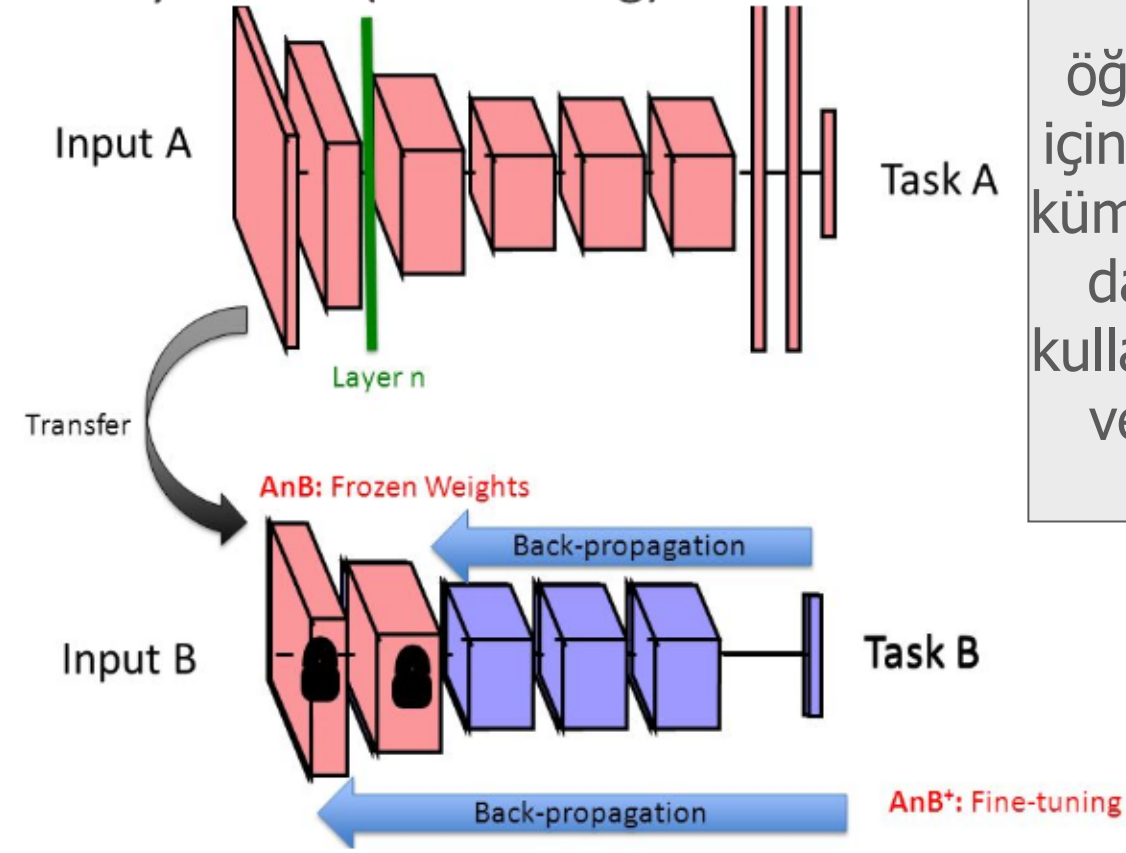
Epoch



ELTME-ERKEN DURDURMA-GÜRÜLTÜ EKLEME

Parametre Hesaplama Yüğü

Model Uyarlama (Finetuning)



Transfer öğrenme ile aslında genellikle yapay öğrenme modelleri ilk katmanlarda hemen her verinin basit özelliklerini çıkarıp öğrenmesinden dolayı ve hemen her basit özellik birbirine benzediği için bu kısımdaki bir başka veride öğrenilmiş parametreler ile yeni veri kümesinde eğitime başlayarak daha az katman için öğrenme sürecinde daha az işlem yükü olacağından dolayı daha hızlı çalışan bir model kullanıyor olursunuz. Özellikle elinizdeki veri kümesi kısıtlıysa büyük bir veri setinde eğitilmiş modelden transfer öğrenme yapıldığında test başarımında hatrı sayılır bir başarımlar yükselmesi olur.

ÇÖZÜM: TRANSFER ÖĞRENME, MULTITASK (ÇOKLU GÖREV) ÖĞRENME



@ayyucekizrak

www.ayyucekizrak.com