



الحمد لله رب العالمين



# An overview of the supervised machine learning methods

مرور کلی بر روش‌های یادگیری ماشین نظارت شده



Vladimir Nasteski

PhD · PhD at University "St. Kliment Ohridski" - Bitola

نویسنده: ولادیمیر ناستسکی

تاریخ: ۱۴۰۴/۰۹/۰۴



Abolfazl ghasemi



Mohammad reza shahbazi



Matin ahmadi

ارائه دهنده: گان





## چکیده | Abstract

- یادگیری نظارت شده یکی از حوزه‌های پراکتیویت در یادگیری ماشین است
- ✓ قابلیت اصلی : استفاده از داده‌های آموزشی برچسب دار (labeled training data).
- این مقاله مروری بر روش‌های بنیادین یادگیری نظارت شده ارائه می‌دهد.
- ✓ هدف : ارائه دید کلی از فناوری‌های یادگیری ماشین و کاربردهای عملی آن‌ها
- کلمات کلیدی: طبقه‌بندی، یادگیری نظارت شده، یادگیری ماشین



## • یادگیری ماشین چیست؟

- یک حوزه گسترده در فناوری اطلاعات، آمار، احتمال، هوش مصنوعی، روانشناسی، نوروپریولوژی و سایر رشته ها.
- حل مسائل از طریق ساخت مدلی که نمایانگر مجموعه داده انتخاب شده است.
- تحول از آموزش کامپیوترها برای شبیه سازی مغز انسان به تولید نظریه های محاسباتی آماری از فرآیندهای یادگیری.

## • هدف اصلی یادگیری ماشین:

- ایجاد الگوریتم هایی که به کامپیوتر اجازه یادگیری می دهند.
- یادگیری = فرآیند یافتن الگوهای آماری یا سایر الگوهای داده ها.



تحول یادگیری ماشین در دهه اخیر :

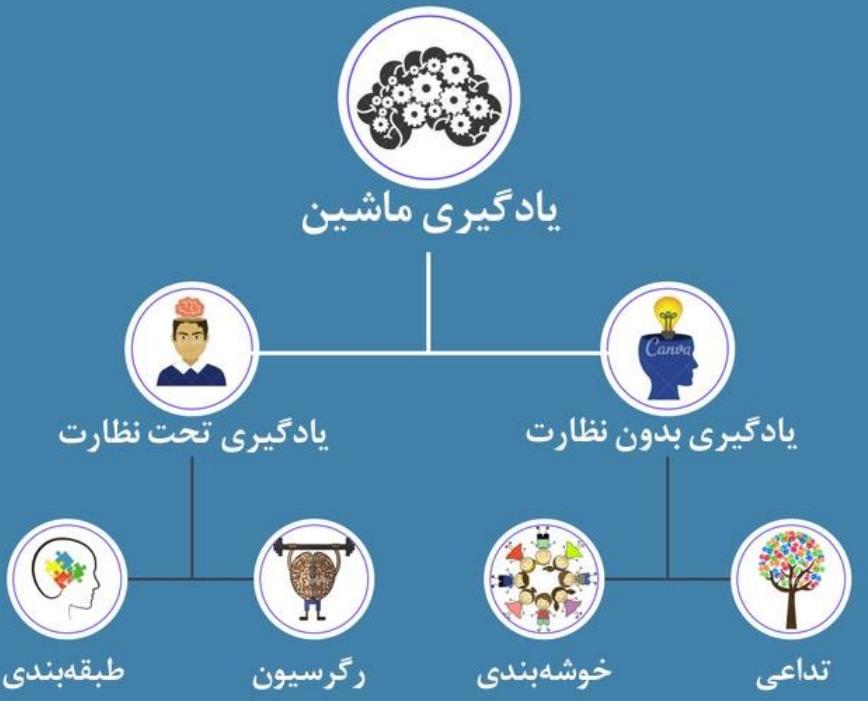
- با توسعه فناوری های محاسباتی در حوزه داده های بزرگ (Big Data)
- الگوریتم ها به طور قابل توجهی توسعه، به روزرسانی و بهبود یافته اند.
- قابلیت اعمال خودکار محاسبات ریاضی پیچیده به داده های حجمی با سرعت بسیار بالا

## برنامه نویسی تطبیقی :

- 1- توانایی شناسایی الگوها
- 2- استخراج اطلاعات جدید از داده ها
- 3- یادگیری از تجربه
- 4- بهینه سازی دقت و کارایی پردازش و خروجی

✓ کاربردهای گسترده: پردازش داده های چند بعدی در حوزه های مختلف

# مقدمه | Introduction



انواع الگوریتم های یادگیری ماشین :

## ۱. یادگیری ناظارت شده (**Supervised Learning**)

تولید تابعی که ورودیها را به خروجی‌های مطلوب نگاشت می‌کند

## ۲. یادگیری بدون نظارت (**Unsupervised Learning**)

مدل‌سازی مجموعه‌ای از ورودی‌ها بدون داشتن نمونه‌های برچسب دار

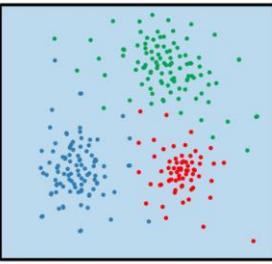
## ۳. یادگیری نیمه ناظارت شده (**Semi-supervised Learning**)

ترکیب نمونه‌های برچسب دار و بدون برچسب

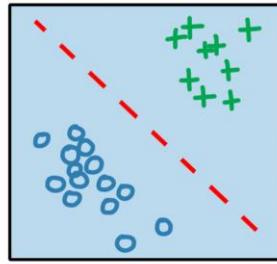


## یادگیری ماشین

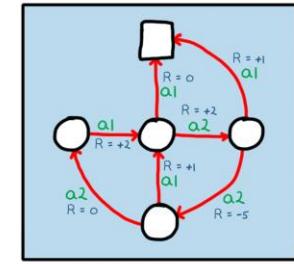
### یادگیری غیرنظراتی



### یادگیری نظارتی



### یادگیری تقویتی



## 4. یادگیری تقویتی (Reinforcement Learning)

یادگیری سیاست عمل در پاسخ به مشاهدات محیط

## 5. یادگیری انتقالی (Transduction)

پیش‌بینی خروجی‌های جدید بر اساس داده‌های آموزشی و ورودی‌های جدید

## 6. یادگیری برای یادگیری (Learning to Learn)

یادگیری سوگیری استقرایی خود بر اساس تجربیات قبلی



## ✓ زیرمجموعه های اصلی یادگیری نظارت شده:

مدل های طبقه بندی (Classifiers)

مدل های رگرسیون (Regression Models)

## ۰ تفاوت کلیدی یادگیری نظارت شده و بدون نظارت: یادگیری نظارت شده :

۱. کلاس ها از پیش تعیین شده اند.
۲. با مجموعه ای محدود و تعریف شده توسط انسان ساخته می شوند.
۳. الگوریتم باید الگوهای را بیابد و مدل های ریاضی بسازد.
۴. ارزیابی بر اساس توانایی پیش بینی

## یادگیری بدون نظارت:

۱. کلاس ها ارائه نمی شوند.
۲. الگوریتم به طور خودکار برچسب های طبقه بندی را توسعه می دهد.
۳. جستجوی شباهت بین قطعات داده برای دسته بندی
۴. ایجاد خوشه ها (Clusters)



# کارهای مرتبط (Related Work)

## تحقیقات پیشین در این حوزه:

Caruana & Niculescu-Mizil (2006):

مقایسه تجربی بزرگ مقیاس بین 10 روش یادگیری نظارت شده شامل SVM ها، شبکه‌های عصبی، رگرسیون لجستیک، (Naive Bayes) و ...

Akritidis & Bozanis (2013):

معرفی الگوریتمی که چندین پارامتر و متا داده‌های مقالات تحقیقاتی را ترکیب می‌کند.

Khan et al. (2010):

مروری بر تکنیک‌ها و روش‌های طبقه‌بندی اسناد متنی

Kotsiantis (2007):

توصیف تکنیک‌های مختلف طبقه‌بندی یادگیری ماشین نظارت شده.

Singh et al. (2016):

بحث درباره کارایی الگوریتم‌های نظارت شده از لحاظ دقیق، سرعت یادگیری، پیچیدگی و خطر بیش برآش (overfitting).

# زمینه یادگیری نظارت شده



فرآیند یادگیری نظارت شده:

دو مرحله اصلی:

• آموزش (Training)

نمونه‌های داده‌های آموزشی به عنوان ورودی گرفته می‌شوند و  
ویژگی‌ها توسط الگوریتم یادگیری استخراج شده و مدل یادگیری  
ساخته می‌شود.

• آزمون (Testing)

مدل یادگیری از موتور اجرا برای پیش‌بینی برای داده‌های آزمون  
یا تولید استفاده می‌کند.

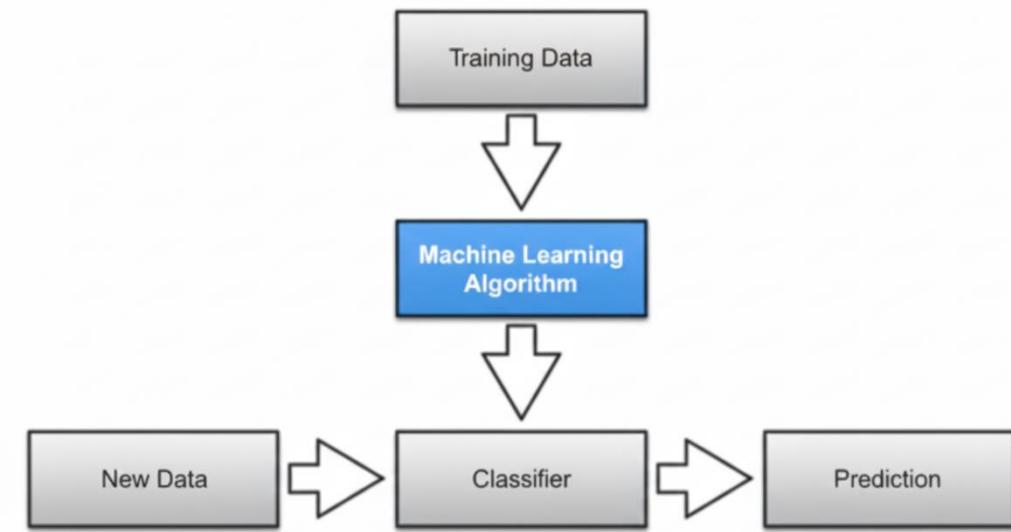


Figure 1: Supervised learning process

# زمینه یادگیری نظارت شده

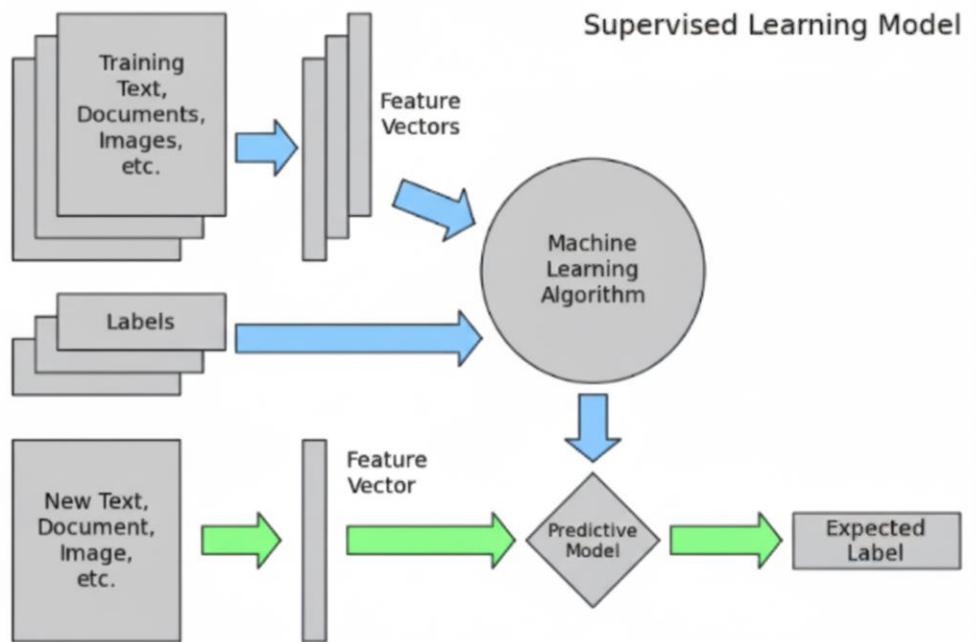


Figure 2: Supervised Learning Model

## مشخصات کلیدی یادگیری نظارت شده:

- ✓ رایج ترین تکنیک در مسائل طبقه بندی
- ✓ هدف: آموزش ماشین برای یادگیری سیستم طبقه بندی که ما ایجاد کردہ‌ایم.
- ✓ اغلب احتمال ورودی‌ها را تعریف نشده باقی می‌گذارد، در حالی که خروجی مورد انتظار شناخته شده است.
- ✓ مجموعه داده شامل ویژگی‌ها و برچسب‌هاست.

► وظیفه اصلی: ساخت برآوردگری (Estimator) که بتواند برچسب یک شیء را با توجه به مجموعه ویژگی‌ها پیش بینی کند.

- الگوریتم یادگیری ورودی‌ها را خروجی‌های صحیح دریافت می‌کند و با مقایسه خروجی صحیح، خطاهای خود را پیدا و مدل را اصلاح می‌کند.

## ❖ انواع مسائل یادگیری نظارت شده:

□ طبقه بندی (Classification) : برچسبها مقادیر گسته دارند. (مقادیر کاتگوریک)

□ رگرسیون (Regression) : برچسبها مقادیر پیوسته دارند (مقادیر اعشاری)

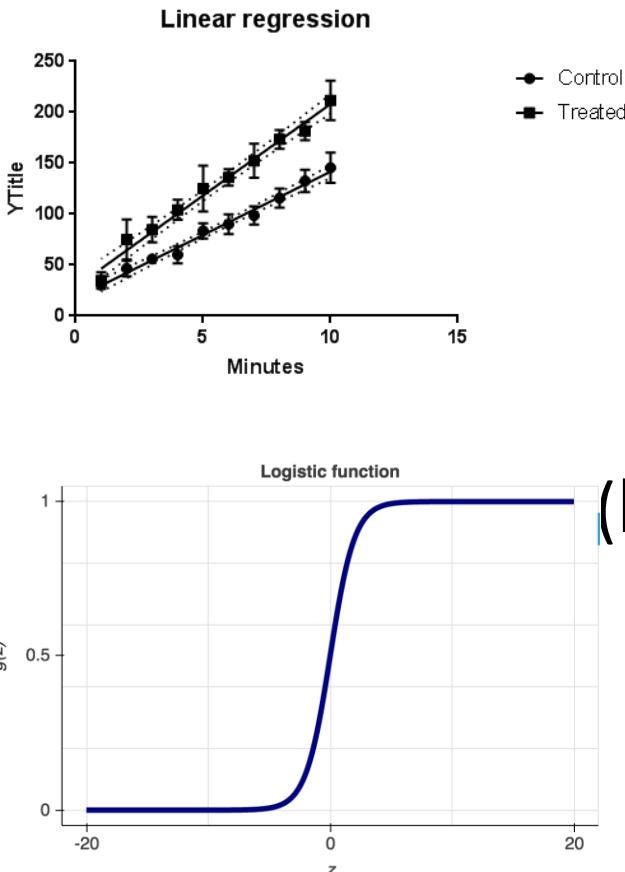
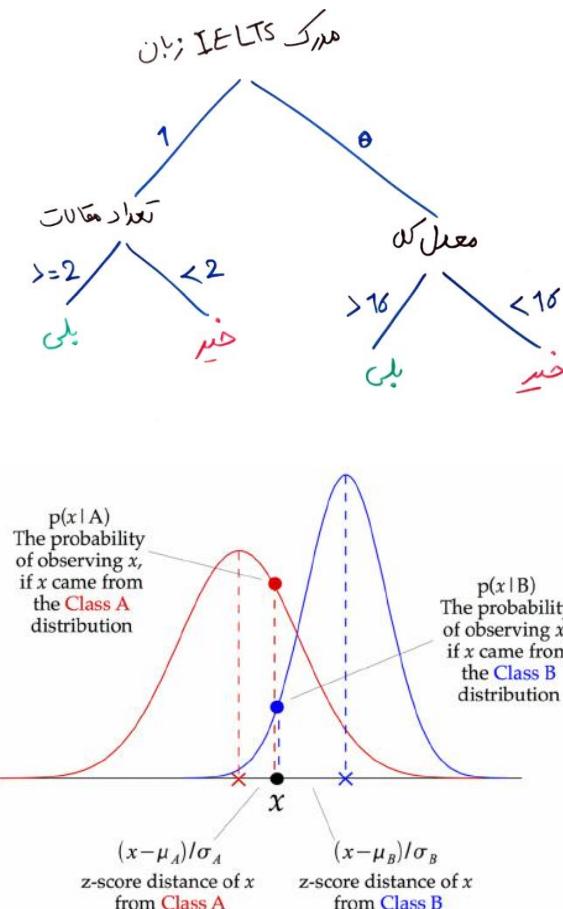
## ❖ کاربردهای عملی:

- پیش بینی رویداد های آینده بر اساس داده های تاریخی
- مثال: پیش بینی گونه های گل ایریس بر اساس اندازه گیری های گل

## ❖ الگوریتم های رایج:

1. شبکه های عصبی
2. ماشین های بردار پشتیبان
3. درختان تصمیم
4. رگرسیون

# الگوریتم های یادگیری نظارت شده



## • الگوریتم های اصلی مورد بررسی :

1. درخت تصمیم (Decision Trees)
2. رگرسیون خطی (Linear Regression)
3. نایو بیز (Naive Bayes)
4. رگرسیون لجستیک (Logistic Regression)

## • اهداف بررسی :

- درک ساختار اساسی هر الگوریتم
- کاربردهای عملی هر روش
- مزايا و محدودیت های هر تکنیک
- تفاوت های کلیدی بین روش ها

# درخت تصمیم (Decision Trees)

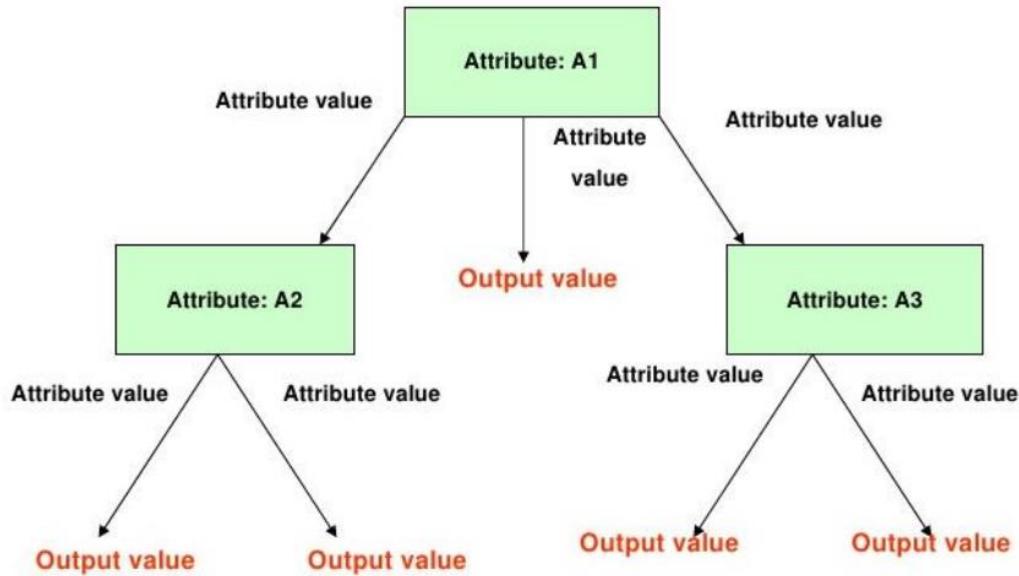


Figure 3: Decision tree example

1. هر گره داخلی فضای نمونه را به دو یا چند زیرفضا تقسیم می‌کند
2. در ساده‌ترین حالت، هر تست یک ویژگی واحد را بررسی می‌کند
3. برای ویژگی‌های عددی، شرط به یک محدوده اشاره دارد
4. هر برگ به یک کلاس اختصاص داده می‌شود که نمایانگر مناسب‌ترین مقدار هدف است

## □ تعریف و مفاهیم:

- یک classifier به صورت بخش بندی بازگشتی فضای نمونه‌ها

## □ ساختار درختی شامل:

- گره ریشه (Root) : بدون یال ورودی
- گره‌های داخلی (Internal Nodes) : تست ویژگی‌ها
- برگ‌ها (Leaves) : برچسب کلاس‌ها

## □ ویژگی‌های کلیدی:

1. هر گره داخلی فضای نمونه را به دو یا چند زیرفضا تقسیم می‌کند
2. در ساده‌ترین حالت، هر تست یک ویژگی واحد را بررسی می‌کند
3. برای ویژگی‌های عددی، شرط به یک محدوده اشاره دارد
4. هر برگ به یک کلاس اختصاص داده می‌شود که نمایانگر مناسب‌ترین مقدار هدف است

# درخت تصمیم (Decision Trees)

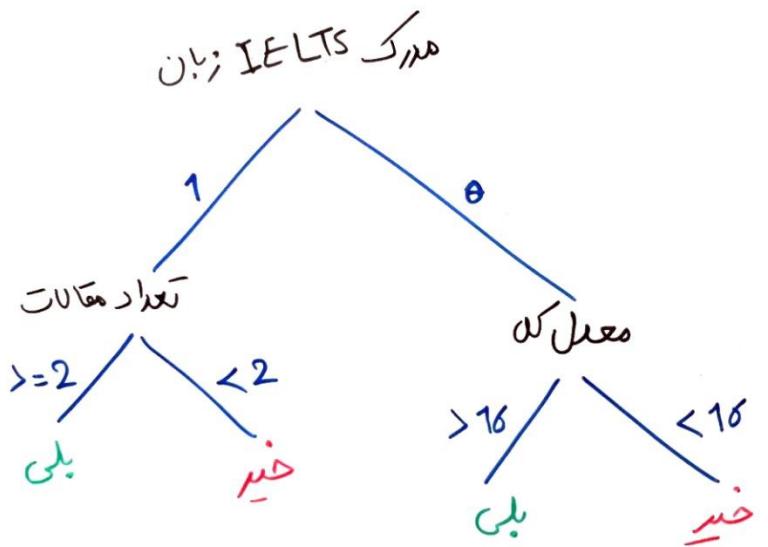


## □ کاربرد عملی درخت تصمیم:

- تحلیلگر می‌تواند پاسخ یک مشتری بالقوه را پیش‌بینی کند
- درک ویژگی‌های رفتاری کل جمعیت مشتریان بالقوه

## □ تفسیر هندسی:

- درختان تصمیم با ویژگی‌های عددی را می‌توان به صورت مجموعه‌ای از صفحات هذلولوی تفسیر کرد
- هر صفحه عمود بر یکی از محورهاست



## □ ترجیحات تصمیم‌گیران:

- درختان تصمیم کم پیچیده‌تر ترجیح داده می‌شوند
- دلیل: جامع تر و قابل فهم تر هستند

## □ چالش اصلی: تمایل به بیش برازش (Overfitting)

# رگرسیون خطی (Linear Regression)

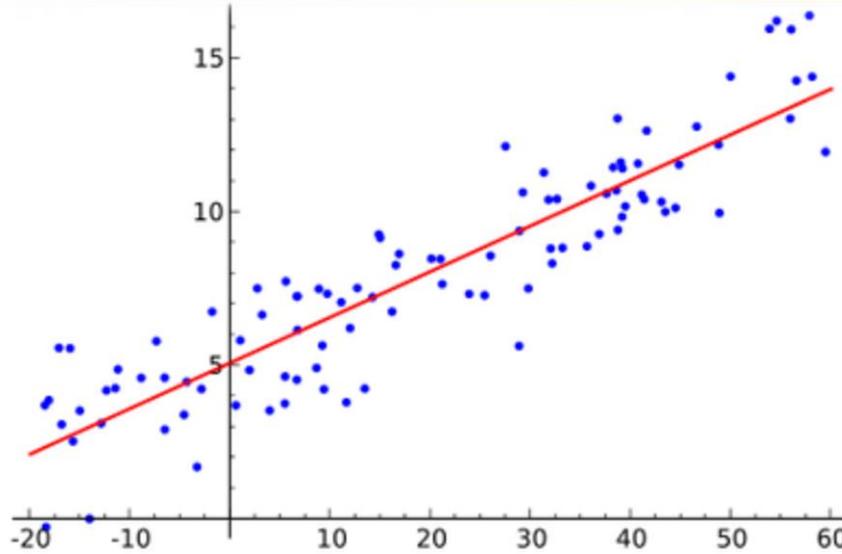


Figure 4: Visual representation of the linear regression

- اهداف و مفاهیم اساسی:

- یافتن روابط و وابستگی‌ها بین متغیرها

- مدل‌سازی رابطه بین:

- یک متغیر وابسته پیوسته ( $y$ )

- یک یا چند متغیر توضیحی ( $X$ )

- حوزه کاربرد:

1. پیش‌بینی یک متغیر هدف پیوسته

2. در مقابل طبقه بندی که یک برچسب از مجموعه محدود را پیش‌بینی می‌کند

- فرم کلی برای رگرسیون چندگانه:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + e$$



# رگرسیون خطی (Linear Regression)

□ ویژگی‌های کلیدی :

جزء روش‌های یادگیری نظارت شده است.

فرآیند آموزش:

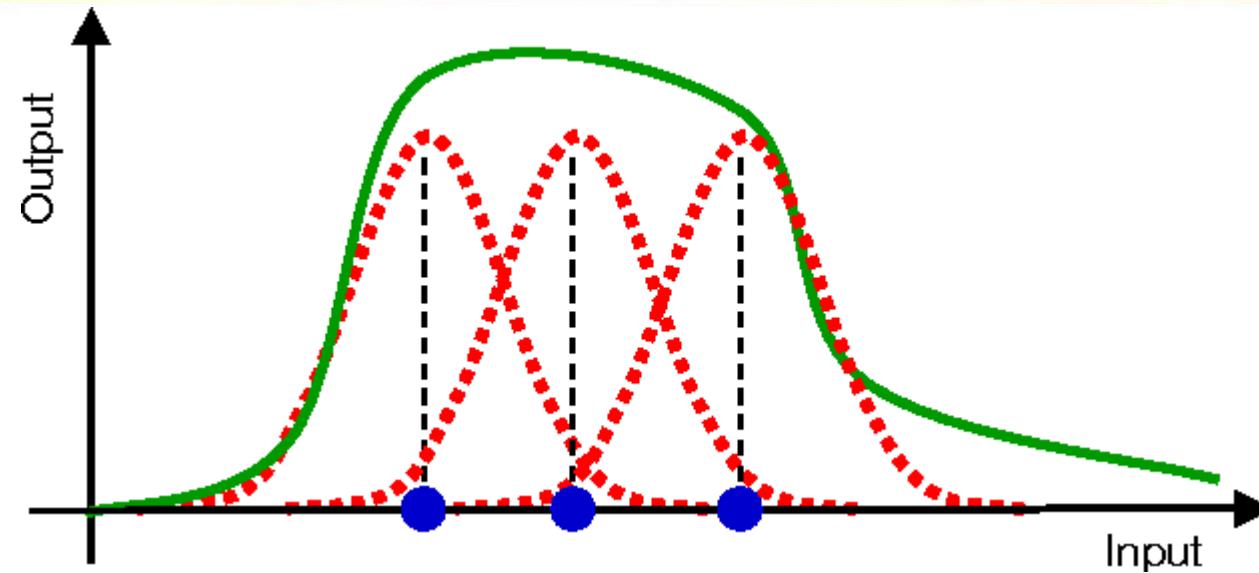
1. مدل بر روی داده‌های برچسبدار (داده‌های آموزشی) آموزش می‌بیند
2. سپس برای پیش‌بینی برچسب‌ها برای داده‌های بدون برچسب (داده‌های آزمون) استفاده می‌شود.

□ فرآیند بهینه‌سازی:

- مدل (خط قرمز) با استفاده از داده‌های آموزشی (نقاط آبی) محاسبه می‌شود
- هر نقطه دارای یک برچسب شناخته شده (محور y) است
- هدف: برازش (Fit) نقاط با کمینه کردن مقدار تابع اتلاف انتخاب شده
- پس از آموزش، مدل می‌تواند برای پیش‌بینی برچسب‌های ناشناخته استفاده شود

✓ کاربردهای عملی: 1. پیش‌بینی قیمت‌ها 2. تحلیل روندها 3. پیش‌بینی مالی

# نایو بیز (Naive Bayes)



## ■ مبانی نظری:

- یک روش یادگیری نظارت شده است
- روش آماری برای طبقه بندی
- بر پایه قضیه بیز عمل می کند

## ■ اهداف اصلی:

- حل مسائل پیش‌بینی
- ترکیب داده‌های مشاهده شده
- ارائه چشم اندازی برای درک و ارزیابی الگوریتم‌های یادگیری
- محاسبه احتمالات صریح برای فرضیه‌ها
- مقاومت در برابر نویز در داده‌های ورودی

## ■ معادله اصلی:

- برای دو متغیر  $x_1, x_2$  و با استفاده از قضیه بیز:

$$P(x_1, x_2) = P(x_1|x_2)P(x_2)$$



# نایو بیز (Naive Bayes)

## ▪ فرضیه کلیدی ناویه بیز:

- استقلال شرطی ویژگی ها نسبت به کلاس
- برای متغیرهای  $x_1, x_2, \dots, x_n$  شرطی نسبت به متغیر دیگری مانند  $C$  :
- که در آن از ۱ تا  $n$  متغیر است.

$$P(x|C) = \prod P(x_i|C)$$

## □ مزایای اصلی:

- محاسبه احتمالات صریح برای هر کلاس
- کارایی بالا در مسائل با ابعاد بالا
- مقاومت نسبی در برابر نویز
- پیاده‌سازی ساده

✓ کاربردهای عملی: 1. تشخیص ایمیل‌های اسپم 2. تحلیل احساسات 3. متن‌کاوی 4. طبقه‌بندی سند‌ها

❖ محدودیت اصلی: فرض استقلال که در عمل همیشه برقرار نیست

# رگرسیون لجستیک (Logistic Regression)

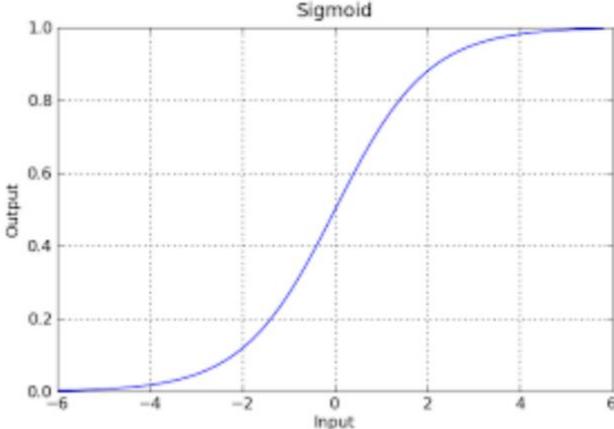


Figure 5: Visual representation of the Logistic Function

## ❖ مقایسه با نایو بیز:

- هر دو با استخراج مجموعه‌ای از ویژگی‌های وزن دار از ورودی کار می‌کنند
- هر دو از لگاریتم‌ها استفاده کرده و به صورت خطی ترکیب می‌کنند
- تفاوت کلیدی: رگرسیون لجستیک یک classifier تمایزی (Discriminative) است در حالی که نایو بیز یک classifier تولیدی (Generative) است.

## ❖ تعریف و مفاهیم:

- نوعی رگرسیون که احتمال وقوع یک رویداد را با برآذش (Fit) داده‌ها با یک تابع لجستیک پیش‌بینی می‌کند
- از چندین متغیر پیش‌بین استفاده می‌کند که می‌تواند عددی یا کاتگوریکال باشد
- تابع فرضیه: که در آن  $g$  تابع سیگموئید است

$$h\theta(x) = g(\theta^T x)$$



# رگرسیون لجستیک (Logistic Regression)

○تابع سیگموئید:

$$g(z) = 1 / (1 + e^{-z})$$

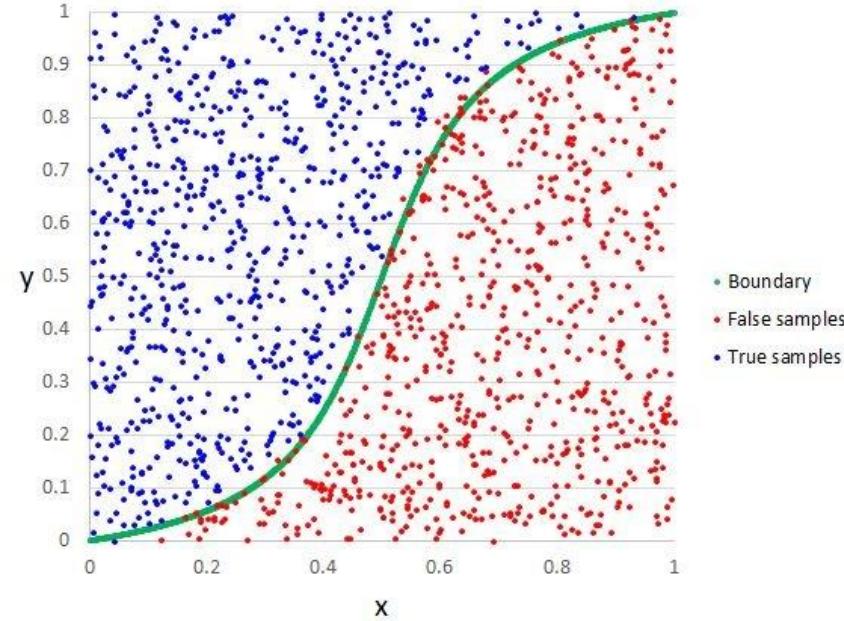
□ ویژگی‌های خاص:

- مقادیر را در بازه  $[0, 1]$  نگه می‌دارد
- ایده‌آل برای پیش‌بینی احتمالات

○تابع هزینه:

$$J(\theta) = \frac{1}{m} \sum [-y(i) \log(h\theta(x(i))) - (1-y(i)) \log(1-h\theta(x(i)))]$$

که در آن  $m$  تعداد نمونه هاست



# رگرسیون لجستیک (Logistic Regression)

## ❖ بهینه سازی (Optimization Process)

- برای یافتن حداقل تابع هزینه، از توابع بهینه سازی مانند **fmin\_bfgs** استفاده می شود.
- این توابع بهترین پارامترها  $(\theta)$  را برای تابع هزینه رگرسیون لجستیک پیدا می کنند.
- پارامترهای نهایی برای رسم مرز تصمیم گیری برای داده های آموزشی استفاده می شوند.

✓ کاربردهای عملی: طبقه بندی دودویی، پیش بینی های باینری



# نتیجه گیری (Conclusion)

## ❖ یافته های کلیدی:

- یادگیری نظارت شده یکی از روش های غالب در یادگیری ماشین است
- روش های نظارت شده در مقایسه با روش های بدون نظارت موفقیت بیشتری داشته اند
- دلیل اصلی: داده های برچسب دار معیارهای واضح تری برای بهینه سازی مدل فراهم می کنند
- مجموعه بزرگی از الگوریتم ها در این حوزه وجود دارد که به طور مداوم توسط دانشمندان داده بهبود می یابند

## ❖ سهم مقاله:

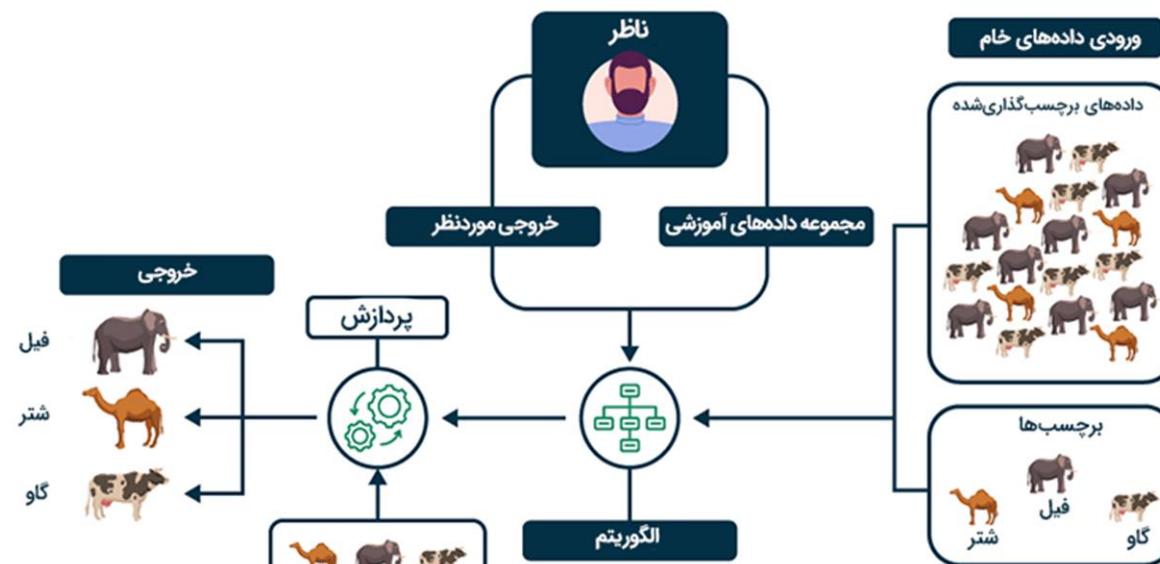
- ارائه دید کلی از چند الگوریتم یادگیری نظارت شده
- توضیح مختصری از فرآیند یادگیری ماشین
- توصیف ساختار اساسی برخی الگوریتم های مختلف یادگیری ماشین

# نتیجه گیری (Conclusion)

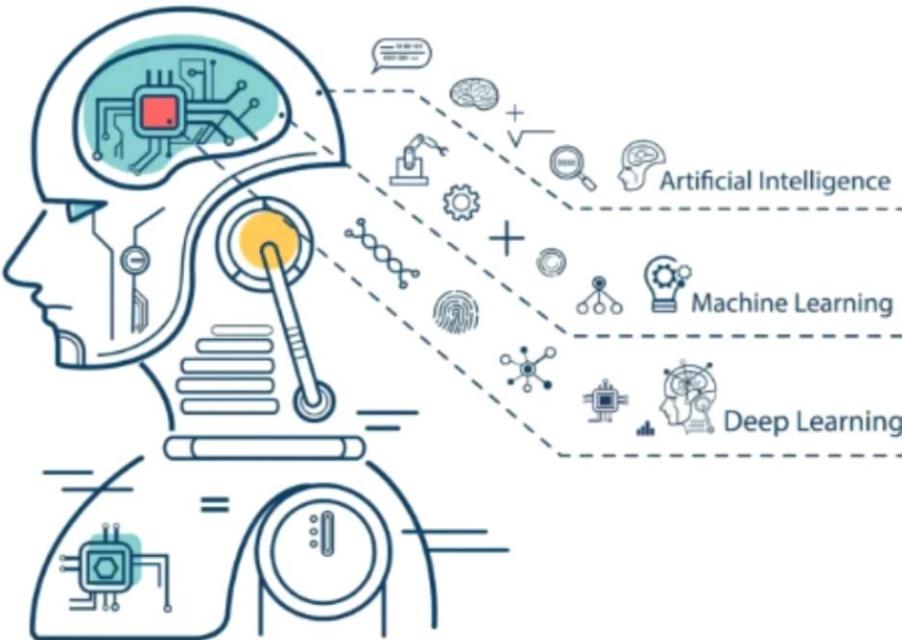
## ✓ پیشرفت‌ها و چشم انداز آینده:

- حوزه یادگیری نظارت شده توجه بسیاری از توسعه‌دهندگان را به خود جلب کرده است
- پیشرفت‌های چشمگیری در دهه اخیر حاصل شده است
- روش‌های یادگیری عملکرد عالی داشته‌اند که در دهه‌های قبل دست یافتنی نبودند

### یادگیری نظارت شده



# نتیجه گیری (Conclusion)



## ❖ فرصت‌های پژوهشی آینده:

1. فضای زیادی برای بهبود روش‌های نظارت شده وجود دارد
2. نیاز به الگوریتم‌های کارآمدتر و قابل تفسیرتر
3. کاربردهای رو به رشد در حوزه‌های: پزشکی و تشخیص پزشکی
4. علوم داده و کلان داده‌ها
5. پردازش زبان طبیعی و متن کاوی
6. تشخیص تصویر و ویدئو
7. تصمیم‌گیری‌های هوشمند

✓ **پیشنهاد نهایی:** ادغام روش‌های نظارت شده با دیگر روش‌های یادگیری ماشین



## منابع کلیدی مقاله:

- Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms.
- Akritidis, L., & Bozanis, P. (2013). A supervised machine learning classification algorithm for research articles.
- Khan, A., et al. (2010). A Review of Machine Learning Algorithms for Text-Documents Classification.
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques.
- Singh, A., et al. (2016). A review of supervised machine learning algorithms.
- Mitchell, T. (2015). Machine Learning.

دريافت مقاله :

- DOI [10.20544/HORIZONS.B.04.1.17.P05](https://doi.org/10.20544/HORIZONS.B.04.1.17.P05)



**JUST DO IT!**

**فقط انجامش بده!**