

# الگوریتم KNN

استاد : طبیبی

نویسنده : پرهام مستجیر

## هوش مصنوعی :

به شاخه‌ای از علوم کامپیوتر اشاره دارد که هدف آن ایجاد سیستم‌ها یا ماشین‌هایی است که می‌توانند وظایفی را انجام ( AI یا Artificial Intelligence هوش مصنوعی ) دهند که معمولاً نیاز به هوش انسانی دارند. این وظایف شامل یادگیری ، استدلال ، حل مسئله، درک زبان طبیعی، تشخیص الگوها و تصمیم‌گیری می‌شود. هوش مصنوعی به سیستم‌ها این توانایی را می‌دهد که از تجربیات گذشته یاد بگیرند، با ورودی‌های جدید سازگار شوند و وظایف پیچیده را به‌طور خودکار انجام دهند.

هوش مصنوعی شامل الگوها و روش‌های مختلفی است که هر کدام برای حل مسائل خاصی طراحی شده‌اند. در ادامه به بررسی برخی از مهمترین این الگوها می‌پردازیم

## تاریخچه هوش مصنوعی :

### بنیانگذاری: 1940-1950

اولین مدل ریاضی نورو ن مصنوعی توسط مک کالوک و پیترز: 1943  
را منتشر کرد "ماشین آلات محاسباتی و هوش" آلن تورینگ مقاله: 1950  
"هوش مصنوعی" تولد رسمی اصطلاح -کنفرانس دارتموث: 1956

### پیشرفتهای اولیه : 1960 دهه

اولین چت بات - (1965) ELIZA توسعه برنامه  
اولین سیستم خبره - (1965) DENDRAL ایجاد سیستم  
(1958) اختراع شبکه عصبی پرسپترون



### اولین زمستان هوش مصنوعی: 1970 دهه

محدودیت‌های سخت‌افزاری و نرم‌افزاری آشکار شد

کاهش بودجه‌های تحقیقاتی

توسعه یافتند (1972) MYCIN اما سیستم‌های خبره مانند

### بازگشت به عرصه: 1980 دهه

توسعه سیستم‌های خبره تجاری

(1986) Backpropagation معرفی الگوریتم

ظهور شبکه‌های عصبی چندلایه

### پیشرفت‌های عملی: 1990 دهه

دیپ بلوگری کاسپاروف را شکست داد: 1997

پیشرفت در یادگیری ماشین

توسعه الگوریتم‌های جدید برای داده‌کاوی

### گسترش کاربردها: 2000 دهه

پیشرفت در پردازش زبان طبیعی  
توسعه سیستمهای توصیه گر  
بهبود الگوریتمهای تشخیص الگو

### انقلاب یادگیری عمیق: 2010 دهه

واتسون در 2011: IBM  
برنده شد Jeopardy!  
تحوّل در 2012: AlexNet  
تشخیص تصویر ایجاد کرد  
را 2016: AlphaGo قهرمان  
شکست داد

### عصر مدل‌های: تاکنون 2020 بزرگ

و مدل‌های زبانی بزرگ GPT ظهور  
پیشرفت در تولید محتوای هوش  
مصنوعی  
متنی، (توسعه مدل‌های چندوجهی  
(تصویری، صوتی)

**Natural Language Processing – NLP** پردازش زبان طبیعی

**Computer Vision** بینایی ماشین

**Speech Processing** پردازش گفتار

**Expert Systems** سیستم‌های خبره

**Fuzzy Logic** منطق فازی

**Artificial Neural Networks - ANN** شبکه‌های عصبی مصنوعی



# یادگیری ماشین (Machine Learning) :

یادگیری ماشین، به عنوان یکی از زیرمجموعه های هوش مصنوعی ، این امکان را به کامپیوترها می دهد که با تحلیل داده ها، الگوها را شناسایی کرده و دانش کسب کنند، بدون آن که نیاز به برنامه ریزی دقیق و دستی داشته باشند. با ارزش بازار جهانی بالغ بر ۲۱,۲ میلیارد دلار در سال ۲۰۲۲ و پیش بینی رشد تا ۲۰۹,۹۱ میلیارد دلار تا سال ۲۰۳۰ یادگیری ماشین به سرعت در حال تبدیل شدن به یکی از صنایع پیشرو در جهان است.

# مفاهیم اصلی در یادگیری ماشین

## ویژگی

ویژگی‌ها یا متغیرها، مشخصه‌های خاصی از داده‌ها هستند که برای آموزش مدل استفاده می‌شوند. به عنوان مثال، در یک مجموعه داده‌ی مربوط به مسکن، ویژگی‌ها می‌توانند شامل متراژ، تعداد اتاق‌ها و محل جغرافیایی باشند.

## الگوریتم‌ها

الگوریتم‌ها مغز یادگیری ماشین هستند. انواع مختلفی از الگوریتم‌ها وجود دارد. رگرسیون، طبقه بندی و خوشه بندی از انواع مهم هستند. انتخاب الگوریتم مناسب برای هر مسئله بسیار مهم است.

## داده‌ها

داده‌ها قلب یادگیری ماشین هستند. داده‌ها می‌توانند ساخت یافته، بدون ساختار یا نیمه ساختار یافته باشند. پیش پردازش داده‌ها شامل پاکسازی، نرمال سازی و کاهش ابعاد است.



## پیش‌پردازش داده‌ها

قبل از آموزش مدل، داده‌ها باید پاک‌سازی و آماده‌سازی شوند. این فرآیند شامل مراحل مختلفی مانند حذف داده‌های گم‌شده، نرمال‌سازی داده‌ها و تبدیل ویژگی‌ها است.

## ارزیابی

پس از آموزش مدل، باید عملکرد آن بر روی داده‌های جدید ارزیابی شود. این کار با استفاده از معیارهایی مانند دقت Accuracy، دقت Precision، یادآوری Recall و F1-Score انجام می‌شود.

## لیبل‌ها

در یادگیری نظارت‌شده، برچسب‌ها مقادیر هدفی هستند که مدل سعی می‌کند آن‌ها را پیش‌بینی کند. به عنوان مثال، در یک مسئله طبقه‌بندی، برچسب‌ها می‌توانند نشان‌دهنده‌ی کلاس‌های مختلفی مانند "گربه" یا "سگ" باشند.

## دسته بندی کلی الگوریتم‌های یادگیری ماشین :

الگوریتم‌های یادگیری ماشین به سه دسته کلی تقسیم می‌شوند

یادگیری نظارت‌شده

یادگیری بدون نظارت

یادگیری تقویتی

سایر دسته‌بندی‌های :

یادگیری نیمه‌نظارت‌شده

یادگیری انتقالی

...

# یادگیری نظارت شده (Supervised Learning)

یادگیری نظارت شده از داده های برچسب دار برای آموزش مدل استفاده می کند. هدف، پیش بینی خروجی برای داده های جدید است. در این روش، مدل با استفاده از داده های آموزشی که دارای برچسب هستند، آموزش داده می شود.

## رگرسیون خطی

برای پیش بینی مقادیر پیوسته استفاده می شود. مثال: پیش بینی قیمت خانه بر اساس متراژ.

## رگرسیون لجستیک

برای طبقه بندی استفاده می شود. مثال: تشخیص ایمیل های اسپم.

## ماشین های بردار پشتیبان (SVM)

برای طبقه بندی و رگرسیون استفاده می شود. مثال: تشخیص چهره.



## طبقه‌بندی (Classification)

پیش‌بینی برچسب‌های گسسته

کاربرد ها :

(خیر/بله) تشخیص ایمیل اسپم

تشخیص بیماری از تصاویر پزشکی

تحلیل احساسات متن

الگوریتم‌ها

رگرسیون لجستیک

(SVM) ماشین بردار پشتیبان

درخت تصمیم

شبکه‌های عصبی

## رگرسیون (Regression)

پیش‌بینی مقادیر پیوسته

کاربرد ها :

پیش‌بینی قیمت مسکن

پیش‌بینی دمای هوا

تخمین زمان تحویل

الگوریتم‌ها

رگرسیون خطی

رگرسیون چندجمله‌ای

درخت تصمیم رگرسیونی

# یادگیری غیر نظارت شده (Unsupervised Learning)

در یادگیری غیرنظارت شده هیچ برچسب یا هدف مشخصی وجود ندارد و مدل باید به طور خودکار اطلاعات مفید را از داده ها استخراج کند.

نمونه های کاربردی

بازاریابی (گروه بندی مشتریان بر اساس رفتار خرید) و زیست شناسی (دسته بندی ژن ها با الگوهای بیان مشابه)

## خوشه بندی K-means

برای گروه بندی داده ها استفاده می شود. مثال: بخش بندی مشتریان.

## تحلیل مولفه های اصلی (PCA)

برای کاهش ابعاد داده ها استفاده می شود.

## دسته‌بندی اصلی الگوریتم‌های یادگیری بدون نظارت :

### خوشه‌بندی :

تقسیم داده‌ها به گروه‌های مشابه بر اساس شباهت‌ها.  
K-Means ،DBSCAN ،Hierarchical Clustering.

### کاهش ابعاد :

کاهش تعداد ویژگی‌ها برای سادسازی داده‌ها و جلوگیری از اضافه‌بار محاسباتی.  
PCA ، t-SNE ، Autoencoders.

### کشف قوانین انجمنی :

یافتن روابط بین متغیرها در داده‌های بزرگ.  
Apriori ،FP-Growth.

### مدلسازی احتمالاتی :

مدلسازی توزیع داده‌ها با روش‌های آماری.  
Gaussian Mixture Models ،Hidden Markov Models.



# یادگیری تقویتی (Reinforcement Learning)

عامل (Agent)	اقدام (Action)
تصمیم گیرنده	عملی که عامل انجام می دهد
محیط (Environment)	حالت (State)
فضایی که عامل در آن عمل می کند	وضعیت فعلی محیط
پاداش (Reward)	
بازخوردی که عامل دریافت می کند	
سیاست (Policy)	
استراتژی انتخاب اقدامات توسط عامل	

یادگیری تقویتی از طریق آزمون و خطا یاد می گیرد. عامل در محیط عمل می کند و پاداش دریافت می کند. هدف، یادگیری سیاستی است که پاداش را به حداکثر می رساند.

## کاربردها :

بازی‌ها (مثل شطرنج یا آتاری)

کنترل ربات

سیستم‌های پیشنهاددهنده

مدیریت منابع در شبکه‌ها

## الگوریتم‌های رایج :

Q-Learning

Deep Q-Network (DQN)

Policy Gradient

Actor-Critic

# یادگیری نیمه نظارت شده (Semi-Supervised Learning)

یادگیری نیمه نظارت شده ترکیبی از یادگیری نظارت شده (Supervised Learning) و یادگیری غیرنظارت شده (Unsupervised Learning) است. در این روش، از داده‌های برچسب‌گذاری شده و بدون برچسب به‌طور همزمان استفاده می‌شود تا مدل بتواند الگوهای موجود در داده‌ها را بهتر یاد بگیرد. این روش زمانی مفید است که داده‌های برچسب‌گذاری شده محدود باشند، اما داده‌های بدون برچسب به وفور در دسترس باشند.

کاربردها :

روش‌های رایج :

(Self-Training) برچسب‌گذاری خودکار  
(Label Propagation) انتشار برچسب  
(Co-Training) یادگیری همزمان

تشخیص چهره  
دسته بندی متن و اسناد  
تشخیص گفتار  
و صداتحلیل احساسات در شبکه‌های اجتماعی



# الگوهای یادگیری نظارت شده (Supervised Learning):

طبقه‌بندی Classification

رگرسیون Regression

## الگوریتم K نزدیکترین همسایه ( KNN ) :

با مقایسه داده جدید با همسایه‌های نزدیک، آن را به کلاسی با بیشترین فراوانی تخصیص می‌دهد

### ۱. محاسبه فاصله

داده جدید با تمام داده‌های آموزشی مقایسه می‌شود

معیار: فاصله اقلیدسی

### ۲. انتخاب همسایه‌ها

نقطه با کمترین فاصله انتخاب می‌شوند K

### ۳. تخصیص کلاس

کلاس پرتکرار بین همسایه‌ها به داده جدید نسبت داده می‌شود

# چالش ها در یادگیری ماشین

## کمپرازش (Underfitting)

مدلهایی که بیش از حد ساده هستند ممکن است نتوانند الگوهای موجود در داده ها را به درستی یاد بگیرند

## بیشبرازش (Overfitting)

مدل به خوبی به داده های آموزش می پردازد، اما به خوبی به داده های جدید تعمیم نمی یابد.

## (Data Quality) کیفیت داده ها

داده های گم شده، نویز یا ناسازگاری در داده ها • می تواند عملکرد مدل را به شدت تحت تأثیر قرار دهد

## (Model Selection) انتخاب مدل مناسب

انتخاب مدلی که بیش از حد ساده یا بیش از حد پیچیده باشد می تواند منجر به بیشبرازش (Underfitting) یا کمبرازش (Overfitting) شود