

An overview of the supervised machine learning methods

DOI 10.20544/HORIZONS.B.04.1.17.P05

UDC 004.85.021:519.718

Vladimir Nasteski

Faculty of Information and Communication Technologies, Partizanska bb,
7000 Bitola, Macedonia
vladimir.nasteski@gmail.com



مروی بر روش‌های یادگیری ماشین تحت نظرارت

ارائه دهنده‌گان: ریحانه بختیاری ، فاطمه بابائی

دانشگاه: ملی مهارت دختران زینب کبری

چکیده پژوهش

هدف اصلی

این مقاله به بررسی جامع الگوریتم‌های یادگیری تحت نظارت می‌پردازد و کاربردهای عملی آن‌ها را در حوزه‌های مختلف مقایسه می‌کند.

رویکرد پژوهش

مرور سیستماتیک روش‌های کلیدی شامل درخت تصمیم، رگرسیون خطی، بیز ساده و رگرسیون لجستیک با تأکید بر مزايا، محدودیت‌ها و کاربردهای واقعی هر الگوریتم.



مقدمه: یادگیری ماشین چیست؟

یادگیری ماشین شاخه‌ای از هوش مصنوعی است که به سیستم‌های کامپیووتری قدرت یادگیری از داده‌ها بدون برنامه‌نویسی صریح مندهد. این حوزه به سرعت در حال گسترش است و کاربردهای فراوانی در صنایع مختلف دارد.

یادگیری بدون ناظارت

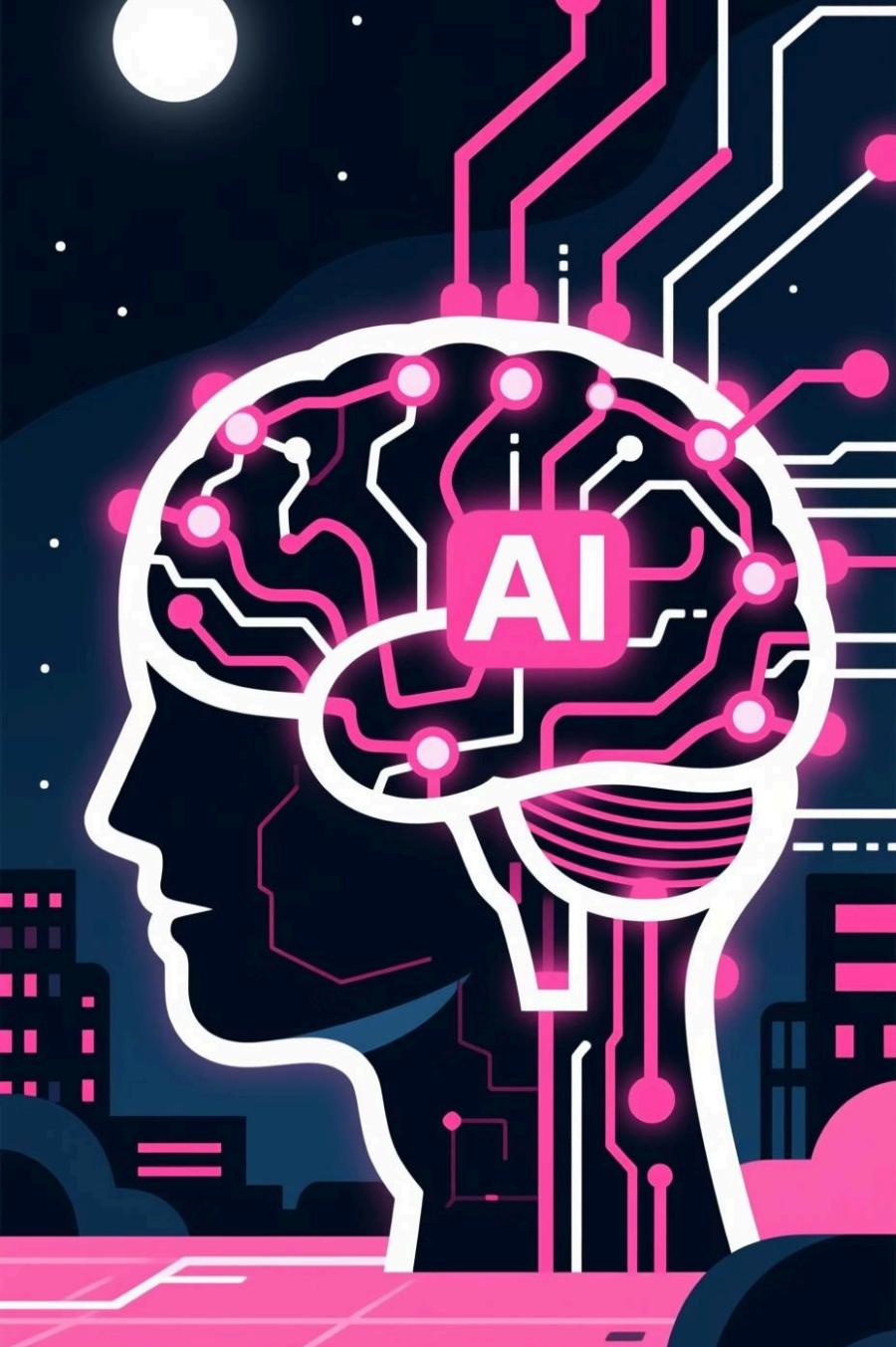
کشف الگوهای و ساختارهای پنهان در داده‌های بدون برچسب

یادگیری تحت نظارت

آموخت با داده‌های برچسبدار و پیش‌بینی خروجی برای ورودی‌های جدید

یادگیری تقویتی

یادگیری از طریق تعامل با محیط و دریافت پاداش یا جریمه



فرآیند یادگیری تحت نظارت

جمع‌آوری داده

گردآوری مجموعه داده‌های برچسب‌دار شامل ویژگی‌های ورودی و خروجی‌های مطلوب

مرحله آموزش

ساخت مدل با استفاده از الگوریتم انتخابی و یادگیری الگوهای موجود در داده‌های آموزشی

مرحله آزمون

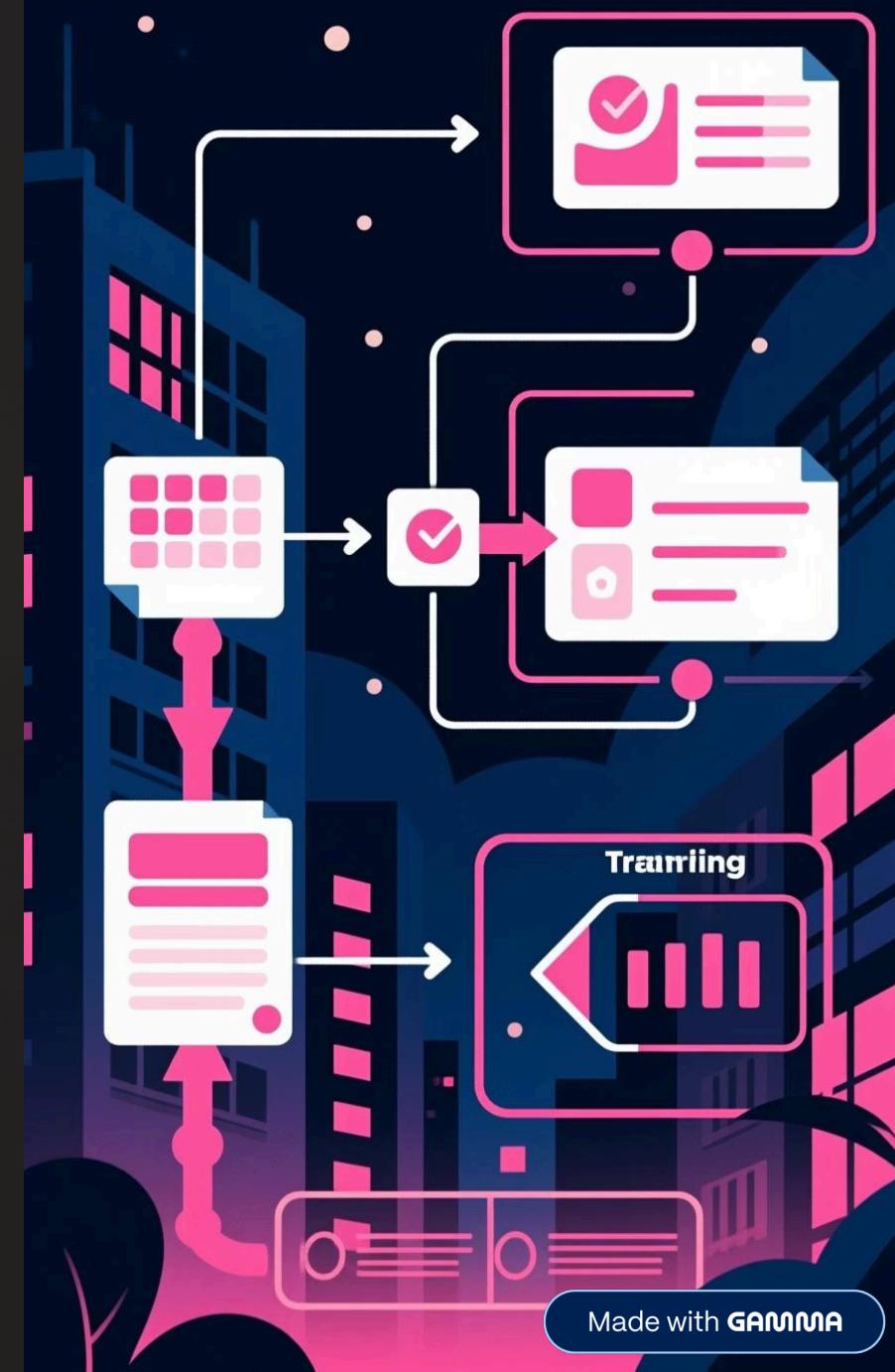
ارزیابی عملکرد مدل روی داده‌های جدید و اندازه‌گیری دقیق پیش‌بینی‌ها

استقرار مدل

به کارگیری مدل نهایی برای پیش‌بینی خروجی برای داده‌های واقعی

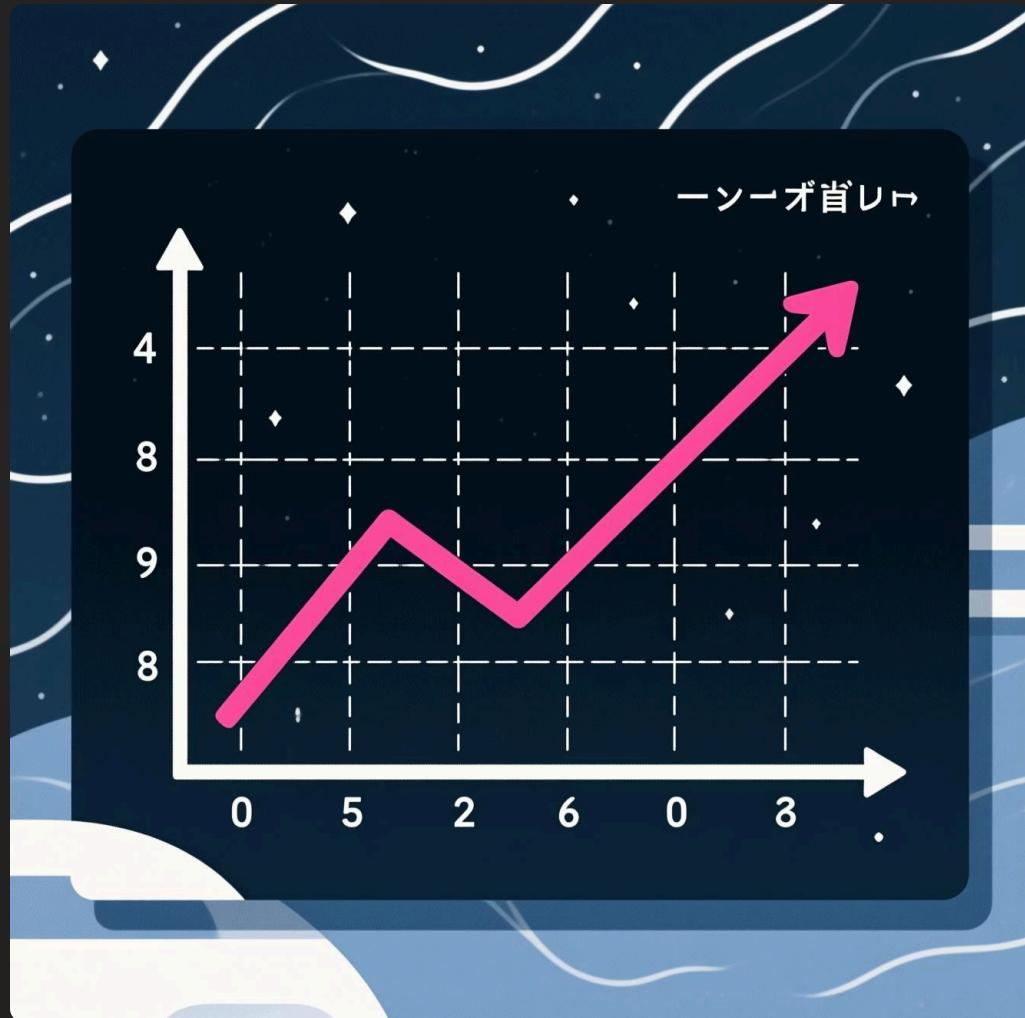
کیفیت و حجم داده‌های آموزشی نقش حیاتی در موفقیت مدل دارد. تفکیک صحیح داده‌ها به مجموعه‌های آموزش و آزمون (معمولًاً ۳۰-۷۰٪ یا ۲۰-۸۰٪) از بیش‌برازش جلوگیری می‌کند.

Supervised Learning Training



رگرسیون در مقابل دسته‌بندی

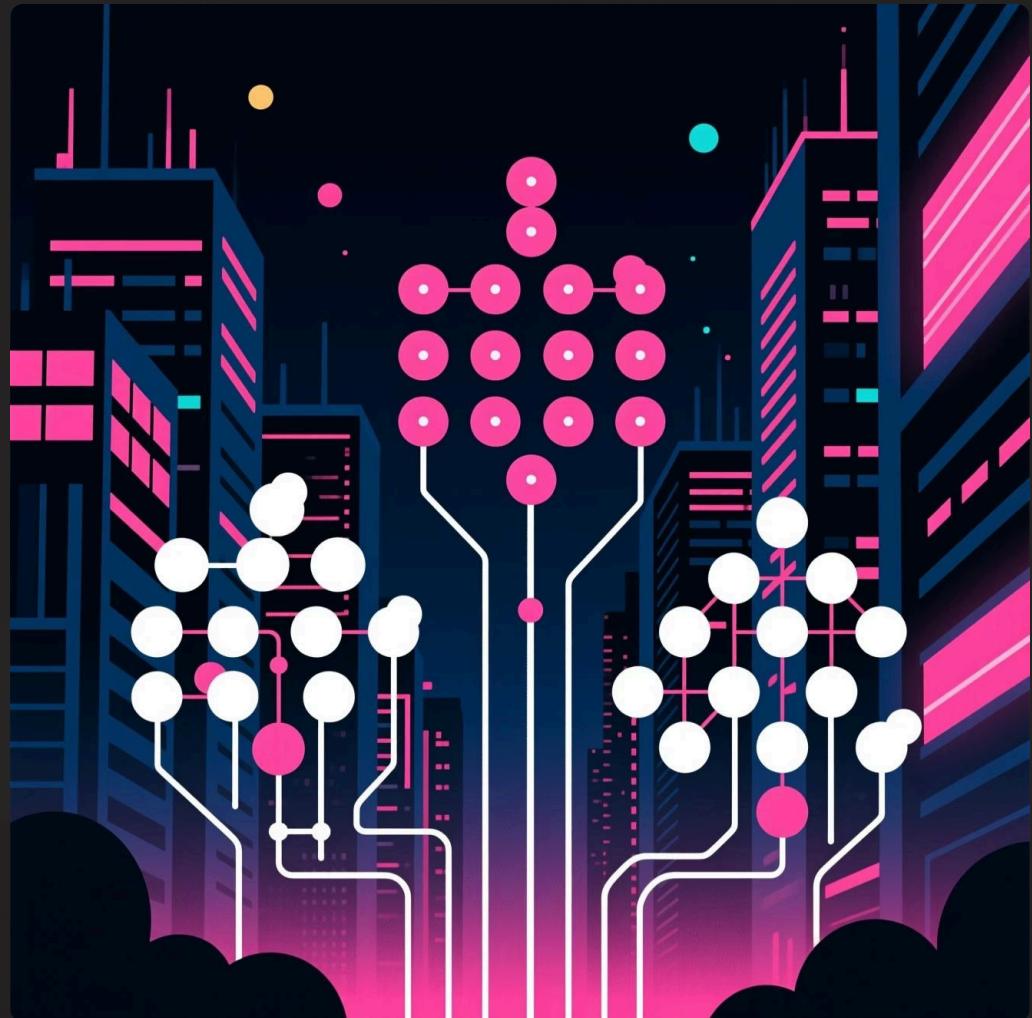
رگرسیون (Regression)



پیش‌بینی مقادیر پیوسته

- خروجی: اعداد حقیقی
- مثال: پیش‌بینی قیمت مسکن، دما، درآمد
- ارزیابی: میانگین خطای مربعات

دسته‌بندی (Classification)



پیش‌بینی برچسب‌های گستته

- خروجی: کلاس یا دسته
- مثال: تشخیص اسپم، تشخیص بیماری، شناسایی چهره
- ارزیابی: دقیق، صحت، فراخوانی



نکته کلیدی: انتخاب بین رگرسیون و دسته‌بندی به ماهیت متغیر هدف (پیوسته یا گستته) بستگی دارد.

درخت‌های تصمیم

درخت‌های تصمیم ساختاری درختی دارند که در آن هر گره داخلی نشان‌دهنده یک آزمون روی ویژگی، هر شاخه نتیجه آزمون، و هر برگ نشان‌دهنده یک برجسب کلاس یا مقدار پیش‌بینی است.

گره‌های داخلی



نقاط تصمیم‌گیری که بر اساس ویژگی‌های ورودی، داده‌ها را به زیرمجموعه‌ها تقسیم می‌کنند

شاخه‌ها



مسیرهای انتخاب که نتایج احتمالی آزمون‌ها را نشان می‌دهند

برگ‌ها



گره‌های انتهایی که تصمیم نهایی یا پیش‌بینی را ارائه می‌دهند

مزایا

- تفسیرپذیری بالا و قابل فهم برای انسان
- نیاز کمتر به پیش‌پردازش داده
- قادر به مدیریت داده‌های عددی و دسته‌ای

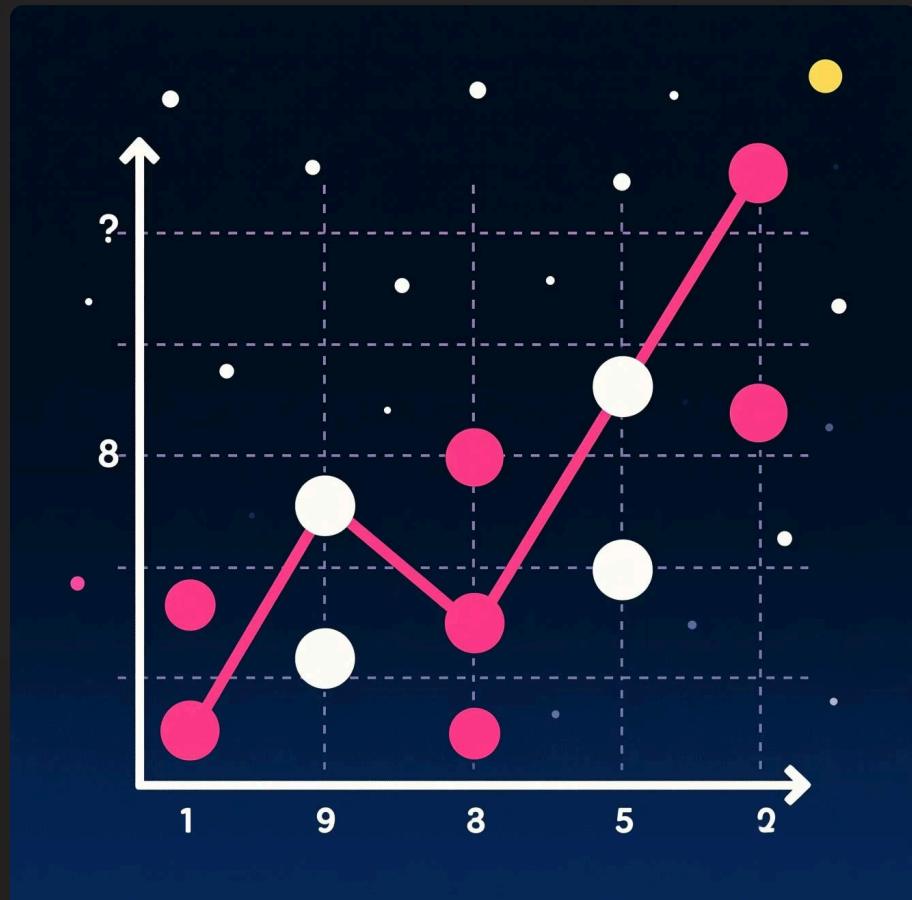
چالش‌ها

- مستعد بیش‌بازش در درخت‌های عمیق
- حساس به تغییرات کوچک در داده
- ممکن است به یهینه محلی برسد



رگرسیون خطی

رگرسیون خطی یکی از ساده‌ترین و پرکاربردترین روش‌های یادگیری ماشین است که رابطه خطی بین متغیرهای ورودی و خروجی را مدل‌سازی می‌کند.



معادله کلی

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

03

ارزیابی مدل

استفاده از معیارهای مانند R^2 و RMSE برای سنجش دقت

02

محاسبه ضرایب

تعیین وزن‌های بهینه برای هر ویژگی ورودی

01

کمینه‌سازی خط

استفاده از روش کمترین مربعات برای یافتن بهترین خط برازش



محدودیت اصلی: فرض خط‌بودن رابطه ممکن است در مسائل پیچیده واقعی صدق نکند و منجر به دقت پایین شود.

طبقه‌بند بیز ساده

الگوریتم بیز ساده بر اساس قاعده بیز عمل می‌کند و با فرض استقلال شرطی بین ویژگی‌ها، احتمال تعلق یک نمونه به هر کلاس را محاسبه می‌کند.



فرض استقلال

فرض می‌شود ویژگی‌ها مستقل از یکدیگر باشند (فرض ساده‌سازی قوی)



قاعده بیز

محاسبه احتمال پسین بر اساس احتمال پیشین و درستنمایی



مقاومت در برابر نویز

عملکرد خوب حتی با وجود داده‌های نویزی یا ناقص

کاربردهای رایج

- فیلتر اسپم ایمیل
- دسته‌بندی متون و اسناد
- تشخیص احساسات
- سیستم‌های توصیه‌گر

ویژگی‌های کلیدی

- سرعت بالا در آموزش و پیش‌بینی
- نیاز به حجم کم داده برای آموزش
- سادگی پیاده‌سازی



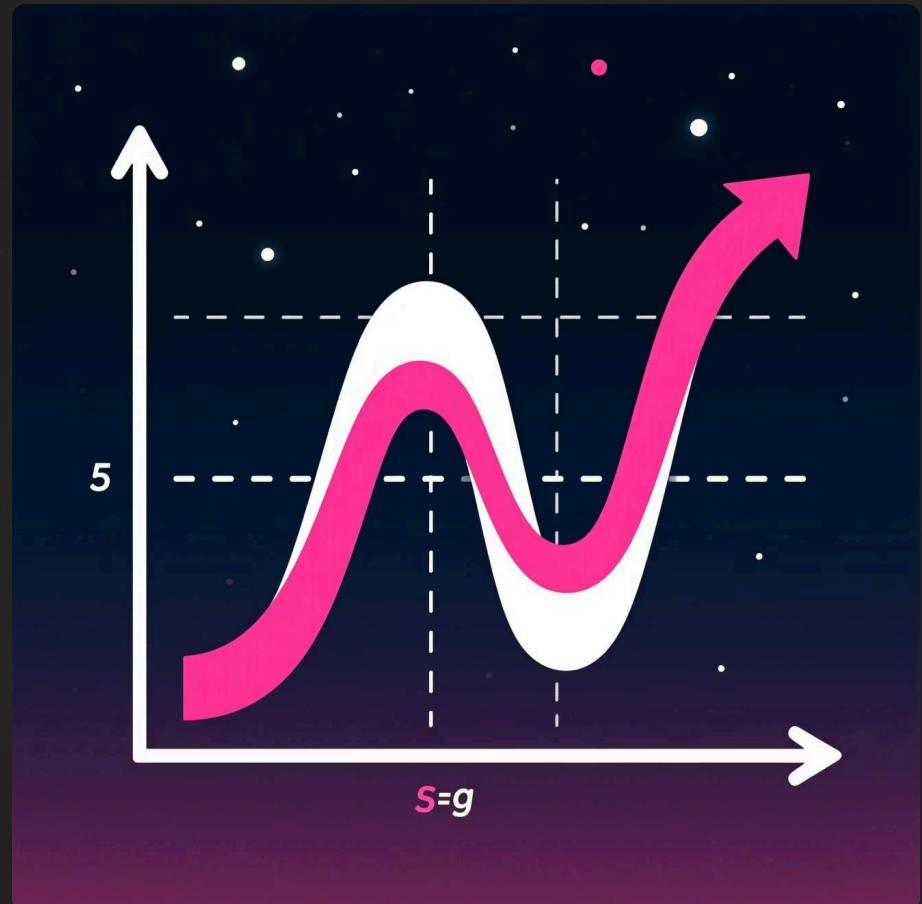
رگرسیون لجستیک

رگرسیون لجستیک یک الگوریتم دسته‌بندی است که با استفاده از تابع سیگموئید، احتمال تعلق یک نمونه به یک کلاس را بین ۰ و ۱ محاسبه می‌کند.

تابع سیگموئید

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

این تابع خروجی خطی را به احتمالی بین ۰ و ۱ تبدیل می‌کند.



رویکرد مولد

Naive Bayes: توزیع احتمال هر کلاس را مدل‌سازی می‌کند (Generative)

رویکرد تمایزدهنده

Logistic Regression: مستقیماً مرز تصمیم بین کلاس‌ها را یاد (Discriminative) می‌گیرد

رگرسیون لجستیک برای دسته‌بندی دودویی بسیار مناسب است و در کاربردهایی مانند تشخیص پزشکی، تحلیل ریسک اعتباری و پیش‌بینی رفتار مشتری استفاده می‌شود.

مقایسه الگوریتم‌ها

الگوریتم	سرعت آموزش	دقت	تفسیرپذیری	حساسیت به نویز
درخت تصمیم	سریع	متوسط تا بالا	بسیار بالا	بالا
رگرسیون خطی	بسیار سریع	متوسط	بالا	متوسط
بیز ساده	بسیار سریع	متوسط	متوسط	پایین
رگرسیون لجستیک	سریع	بالا	بالا	متوسط

انتخاب الگوریتم مناسب به ماهیت مسئله، حجم و کیفیت داده‌ها، نیازهای تفسیرپذیری و محدودیت‌های محاسباتی بستگی دارد. هیچ الگوریتم واحدی برای همه مسائل بهینه نیست.



TONGYI
Qwen

برنامه برای تحلیل مقاله

برنامه
Qwen



Gamma

برنامه ساخت
Power Point

برنامه .app
Gamma