

محمد پور خاکساری

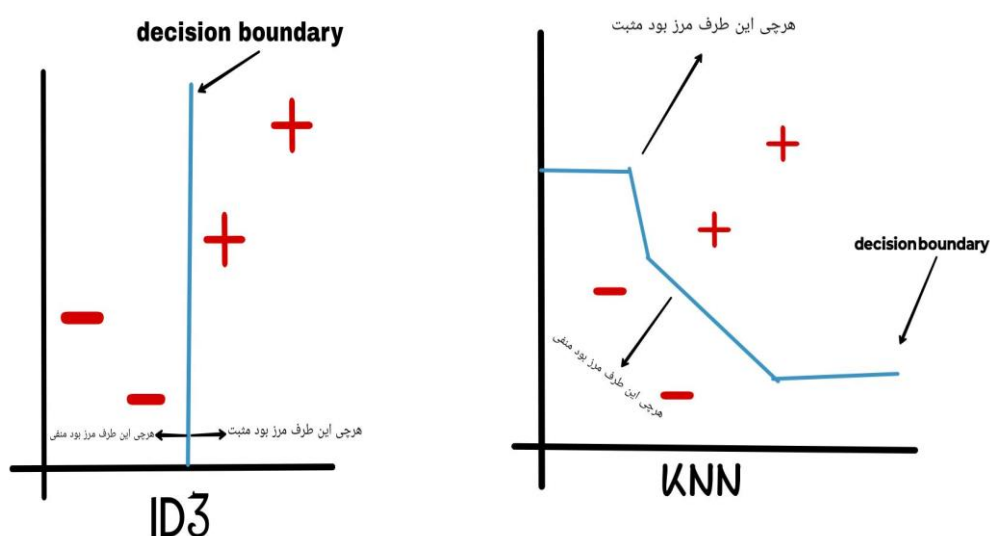
learning = Classification

ما یادگیری را برابر با طبقه‌بندی قرار داده‌ایم:

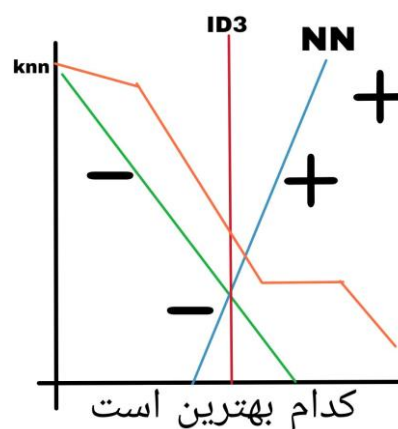
اگر عامل agent بتواند داده‌ها را به درستی تشخیص داده و طبقه‌بندی کند، مثلاً دایره‌ها را از مثلث‌ها تشخیص دهد و مرز بین آن‌ها بکشد،

برچسب‌ها Labels ما قرار است کلاس یا طبقه داده‌ها را مشخص کنند و می‌توانند انواع مختلفی داشته باشند. اگر برچسب‌های ما پیوسته continuous باشند، مسئله ما رگرسیون می‌شود و اگر دودویی Binary، باینری Boolean، یا گسسته discrete باشند، مسئله ما طبقه‌بندی Classification است.

ما به مرز بین طبقات مرز تصمیم‌گیری Decision Boundary می‌گوییم که خطی است که مرز بین دو طبقه را مشخص می‌کند



بر اساس اصل تیغ اوکام Occam's Razor، بهترین راهکار برای حل یک مسئله ساده‌ترین راه است و ساده‌ترین راه برای حل مسئله ما کشیدن یک خط صاف است.



الگوریتم جنگل تصادفی Random Forest تشکیل شده از چند الگوریتم درخت تصمیم‌گیری decision tree که هر کدام نتیجه خودشان را جداگانه به دست می‌آورند تا به الگوریتم بهتری برسیم.

تعریف‌ها:

ترکیب Ensemble: ترکیب چند مدل ضعیف با هم برای به دست آوردن یک مدل قوی‌تر.

جمع‌بندی Bagging: جمع نکات مثبت مدل‌ها برای به دست آوردن طبقه‌بندی درست.

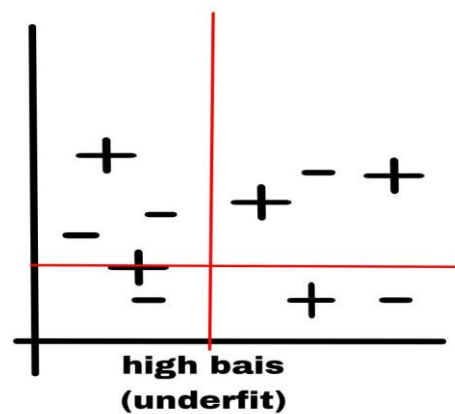
تقویت Boosting: مدل‌ها به صورت مشترک طبقه‌بندی می‌کنند.

پشته‌سازی Stacking: مدل‌ها نتیجه خود را به دست می‌آورند و سپس یک مدل تصمیم می‌گیرد که کدام مدل درست طبقه‌بندی کرده است.

رای‌گیری Voting: مدل‌ها رأی می‌دهند و برجستگی که بیشتر تشخیص داده شد، انتخاب می‌شود.

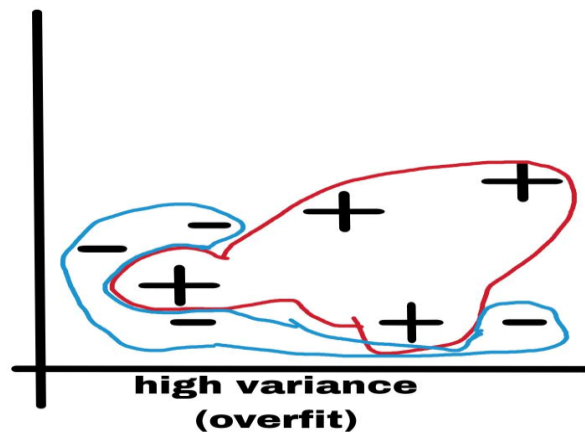
کم‌برازش Underfit: مدل نتوانسته داده‌ها را درست طبقه‌بندی کند و دچار خطای بالا high bias شده است.

خطای بالا High bias: داده‌ها بیش از حد ساده شده هستند و مدل نمی‌تواند درست طبقه‌بندی کند.



بیش‌برازش Overfit: مدل بیش از حد دقیق و درست کار می‌کند و دچار واریانس بالا high variance شده است.

واریانس بالا High variance: داده‌ها به خوبی دسته‌بندی نشده‌اند و مدل درست یاد نگرفته و نمی‌تواند درست طبقه‌بندی کند.



$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

## confusing matrix

		مقدار های واقعی (Actual values)	
		+	-
مقدار های پیشین (Previous values)	+	TP	FP
	-	FN	TN

### TP\_ True positive

مقدار واقعی مثبت بود و الگوریتم هم مثبت پیش بینی کرد

### FP\_ False positive

مقدار واقعی منفی بوده اما الگوریتم مثبت پیش بینی کرد

### FN\_ False negative

مقدار واقعی مثبت بود اما الگوریتم منفی پیش بینی کرد

### TN\_ True negative

مقدار واقعی منفی بوده و الگوریتم هم منفی پیش بینی کرد