

①

(1) در این حالت وقتی باروش هاتیمین نظیر Parzen، توزیع برای ویژگی‌های دگر (دسترس) نشد بدست آورد. روش دیگر نیز آن است که داده‌ها را بدون آن ویژگی برای آموزش مدل استفاده کنیم.

(2) وقتی از Oversampling، Undersampling، روش‌های Cost-Sensitive و روش‌های Ensemble استفاده کرد

(3) وقتی از روش‌های حذف ویژگی استفاده کردو یا Generalization مدل را افزایش داد تا تأثیر ویژگی‌ها در مدل کمتر شود.

(4) وقتی چند ویژگی را حذف کرد و ویژگی که بیشترین correlation با target variable را دارد را نگه داشت

② وقتی از مدل multiple linear regression استفاده کرد:

$$X_1 = \begin{cases} 0 & \text{عمیگرند آکوف} \\ 1 & \text{شده در آنف} \end{cases}$$

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_0$$

حال نرم‌لر معادله رگرسیون را شکلید و دهی:

$$X = \begin{pmatrix} x_1 \\ x_2 \\ 1 \end{pmatrix}$$

$$y = \alpha^T X$$

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_0 \end{pmatrix}$$

$$\Rightarrow \alpha = (X^T X)^{-1} X^T y$$

اگر تعداد داده‌ها زیاد باشد یا ابعاد داده‌ها بالا باشد، استفاده از روش فوق عملاً ناممکن است؛ لذا بایست از روش gradient-descent برای پیدا کردن α استفاده کرد:

$$\alpha := \alpha - \eta \sum_{i=1}^n (y_i - (\alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_0)) \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}$$

\hookrightarrow learning rate

روش فوق، روش Batch GD است؛ پوتران از روش Stochastic GD نیز استفاده کرد:

for $i=1$ to n

$$a =: a - \eta (y_i - a^T x_i) x_i$$

}

که دارای شرایط بسیار کمتری است اما ممکن است در local optimum گیرند.

⑧

• Confusion Matrix

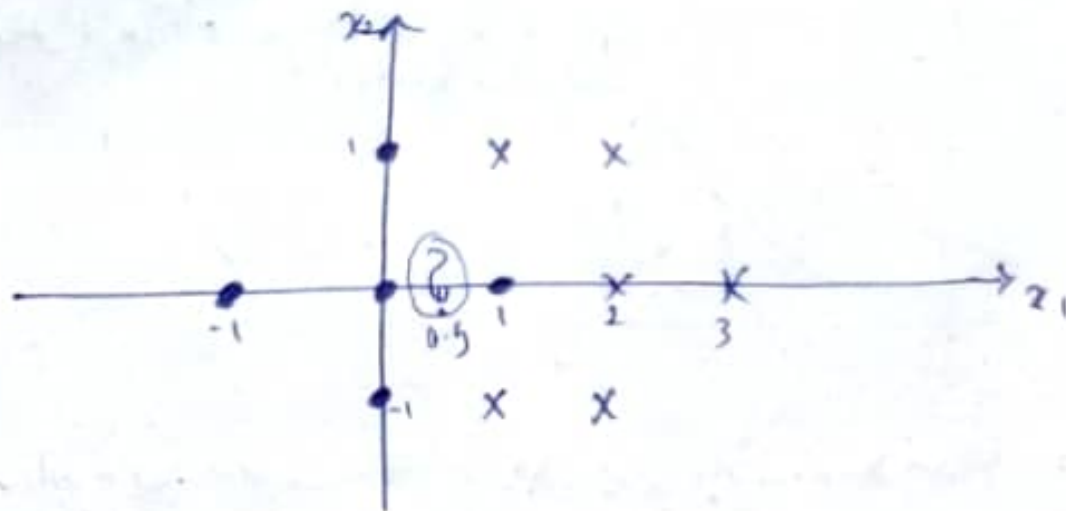
		predicted labels	
		positive	negative
True labels	positive	300	20
	negative	30	200

$$\text{recall} = \frac{TP}{TP+FN} = \frac{300}{300+20} = \frac{300}{320} = 0.9375$$

$$\text{precision} = \frac{TP}{TP+FP} = \frac{300}{300+30} = \frac{300}{330} = 0.9090$$

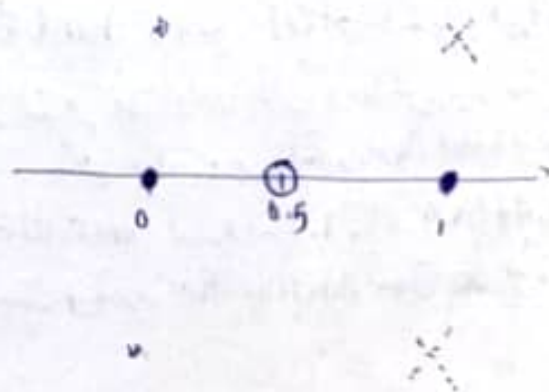
$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{300+200}{300+200+30+20} = \frac{500}{550} = 0.9090$$

$$F_1\text{-Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{1.704375}{1.8465} = 0.9230$$



۱- اقلویسی

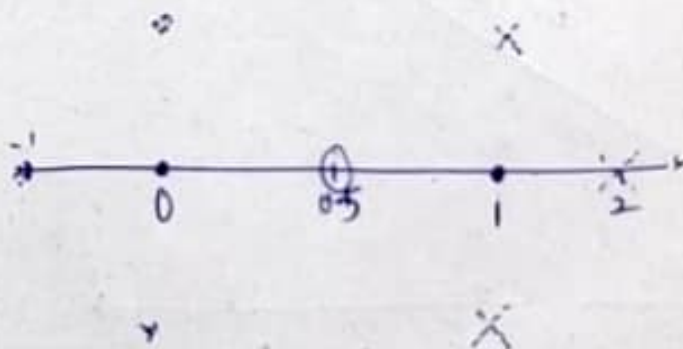
در این حالت 3 نزدیکترین نقاط به صورت زیر اند



نقطه مستقر مرکز از نقاط
تعیین شده و بتواند باشد اما
تأثیر در جواب ندارد
class بیشترین شدن کلاس 2
است.

۲- مسکن

در این حالت 3 نزدیکترین نقاط به صورت زیر اند



این حالت نیز مانند حالت بالاست و نقاط هاسدر و روانه نقطه مستقر باشند اما کلاس بیشترین شده
2 است.

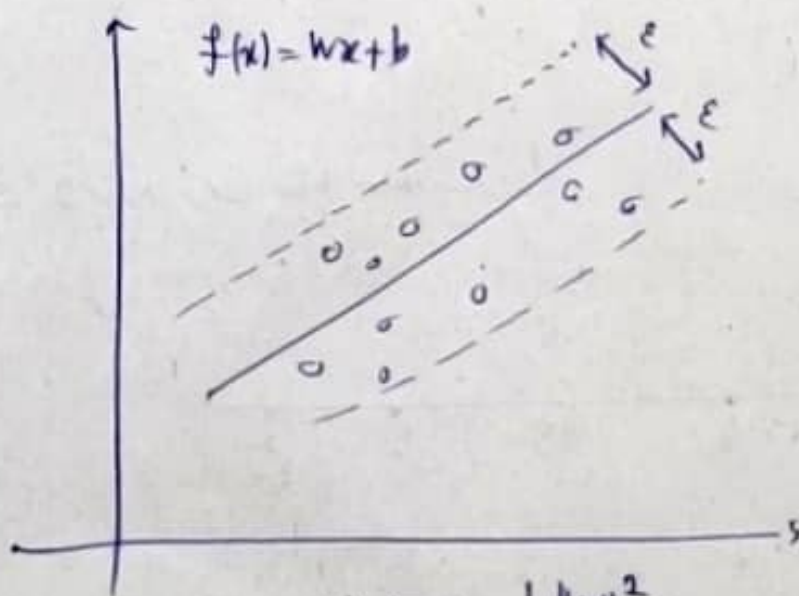
• نقاطی از هر کلاس که به مرز تقسیم نزدیکتر هستند (نسبت به سایر نقاط) $Support\ Vertices$ هستند.



• داده‌هایی که به صورت خطی جدایی پذیر نباشند، توسط SVM به خوبی دسته‌بندی نمی‌شوند.

• کرنل‌ها وظیفه دارند با انجام دادن نگاشت ورود داده ورودی و مپ کردن آن به فضای با ابعاد بالاتر، داده‌ها را با خط جدایی پذیر کنند.

• در $hard\ SVM$ ، هدف آن است که یک ابرصفحه (ایجاد بالا)، جدا شود که $margin$ را بیشینه می‌کند و همه داده‌های آموزش به درستی طبقه‌بندی می‌شوند؛ در این صورت اگر داده‌ها به صورت خطی جدایی پذیر نباشند $hard\ SVM$ سعی می‌کند خط را پیچیده‌تر کند که منجر به $overfitting$ و $Generalization$ کمتر می‌شود. اما در $Soft\ SVM$ ، درجه‌ای از $misclassification$ مجاز است که باعث می‌شود $Generalization$ مولد بیشتر شود و نسبت به نویز مقاوم‌تر شود.



$$\text{minimize: } \frac{1}{2} \|w\|^2$$

$$\text{constraints: } y_i - wx_i - b \leq \epsilon$$

$$wx_i + b - y_i \leq \epsilon$$