SVM 01

Machine Learning — 2019

## Q: Goes for what?

## Topics:

1) Geometric Perceptrons
2) Hard margin classifier
3) Soft margin classifier
4) Hinge loss
5) Comparison of loss functions.
6) Extend the model.

## Perceptron!

1) Initialize
2) Choose "bad" point (random)
3) Normalize it.
4) Update (multiply)
5) Normalize again.
6) Repeat (until no bad points)

## Geometric Problem.

Goal: Find that $\pi$ that maximizes the "margin".

"Margin" maximizing classifier

Linear separability

## (1) Separating hyperplane

$\pi^T \| \pi \| = 1$

Goal: "Maximum Margin"

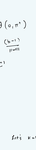
Sol: use the point that barely touches.

$\pi^T = w^T x + b = 0$

$\pi^T = w^T x + b = 1$

$\pi^T = w^T x + b = -1$

$\theta = \frac{w}{\|w\|}$




$\pi^T \cdot w^T x + b = 0$

$w^T x + b = 1$

$w^T x + b = -1$

$\theta = \frac{w}{\|w\|}$

$\theta^T \cdot w^T x + b = 0$

$\theta^T \cdot (w^T x + b) = 0$

$\theta^T \cdot w^T x + \theta^T \cdot b = 0$

$\theta^T \cdot w^T x + \theta^T \cdot b = 1$

$\theta^T \cdot w^T x + \theta^T \cdot b = -1$

$d(\pi, \pi^*) = d(\theta, \pi^*) - d(\theta, \pi)$

$\Rightarrow \frac{|w^T x + b|}{\|w\|} = \frac{|\theta^T x + b|}{\|w\|}$

$\Rightarrow \frac{|w^T x + b|}{\|w\|} = \frac{1}{\|w\|}$

$\pi^T : w^T x + b = +k$

$\pi : w^T x + b = 0$

$\pi^* : w^T x + b = -k$

Both sides

$\pi^T : w^T x + b = k$

$\pi : w^T x + b = 0$

$\pi^* : w^T x + b = -k$

$d = \frac{|k|}{\|w\|} = \frac{|\theta^T x + b|}{\|w\|}$

Either $k = 1$

$\frac{|\theta^T x + b|}{\|w\|} = \frac{1}{\|w\|}$

Either $k = 0$

$\frac{|\theta^T x + b|}{\|w\|} = \frac{|k|}{\|w\|}$

$\theta x_1 + b = x_2 + b$

for code $\boxed{\text{SVM}}$

Fact!

Hinge

$y \in \{-1, 1\}$

$\hat{y} \in \{-1, 1\}$

$\text{Loss}_{\text{hinge}}(y, \hat{y}) = \max(0, 1 - y\hat{y})$

## Hard margin classifier

$\underset{w,b}{\operatorname{argmax}} \left[ \frac{2}{\|w\|} \right]$

s.t. $\left( \underbrace{w^T x_i + b}_{\hat{y}_i} \right) y_i \geq 1$

$\hat{y}_i = w x_i$

$\hat{y}_i \cdot y_i \geq 1$

$\underset{w,b}{\operatorname{argmin}} \left[ \frac{\|w\|^2}{2} \right]$

s.t. $\left( w^T x_i + b \right) y_i \geq 1$

$\underset{w,b}{\operatorname{argmin}} \left[ \frac{\|w\|^2}{2} \right]$

s.t. $\left( \underbrace{w^T x_i + b}_{\hat{y}_i} \right) y_i \geq 1$

$\hat{y}_i \cdot y_i \geq 1$

Q: Why always $y_i \cdot \hat{y}_i \geq 1$?

Assumption: $\hat{y}_i \cdot y_i \geq 1$

$\left( w^T x_i + b \right) y_i \geq 1$

$w^T x_i + b \geq 1$ for positives.

$w^T x_i + b \leq -1$ for negatives.

Hard margin SVM

## Soft margin classifier (aka Hinge loss)

$y_1 \cdot (w^T x_1 + b) \geq 1 - \theta_1$

$\theta_1 \geq 0$ for all "correctly" placed points

$\theta_2 > 0$ for all "incorrectly" placed points

$\theta_i \geq 0$

## Optimization curve

$\underset{w,b}{\operatorname{argmin}} \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \theta_i \right]$

Soft SVM

$\underset{w,b}{\operatorname{argmin}} \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \theta_i \right]$

s.t. $y_i (w^T x_i + b) \geq 1 - \theta_i$

$\theta_i \geq 0$

$\theta_i \longrightarrow$

## For various Datasets.

$C > 0$

$\boxed{C = 10^6}$ → focus more on errors (low).

So, margin will be small

↓

overfit




$\boxed{C = 10^{-6}}$ → Also cannot take too much (low)

↓

Just the margin

↓

Underfit

Underfit

Q: What's the value of C?

→ Hyperparameter tuning

→ CV

→ Cross-Val.

C

$C \uparrow \rightarrow$ overfit

$C \downarrow \rightarrow$ underfit

$\hat{y} > 1 \rightarrow$ correct

$\hat{y} \leq 1 \rightarrow$ wrong

$\boxed{C = 1}$

Default value

$\underset{w,b}{\operatorname{argmin}} \left[ \frac{\|w\|^2}{2} + C \sum_{i=1}^{n} \theta_i \right]$

s.t. $y_i (w^T x_i + b) \geq 1 - \theta_i$

$\theta_i \geq 0$

$\theta_i \longrightarrow$

$\sum_{i=1}^{n} \theta_i$

$\hat{y}_i \cdot y_i$

$\hat{y}_i \cdot y_i \geq 1$

$\hat{y}_i \cdot y_i \leq 1$

## Hinge loss.

$y_i (w^T x_i) \geq 1 - \theta_i$

$\theta_i \geq 1 - y_i (w^T x_i)$

$\theta_i \geq 0$

$\theta_i \geq \max(0, 1 - y_i (w^T x_i))$

$\theta_i = \max(0, 1 - \hat{y}_i \cdot y_i)$

hinge loss

$\theta_i = \max(0, 1 - 0.5)$

$\max(0, -0.5)$

$0$

$\theta_i = \max(0, 1 - 1.5)$

$\max(0, -0.5)$

$0$

$\theta_i = \max(0, 1 - 0.5)$

$\max(0, 0.5)$

$0.5$

$\theta_i = \max(0, 1 - (-0.5))$

$\max(0, 1.5)$

$1.5$

## Various loss functions

$\{-1, 1\}$

$\{-1, 1\}$

$-\left[ y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$

$\hat{y}_i \in \{-1, +1\}$

$\boxed{\text{SVM}}$

$\left[ \log \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i (w^T x_i + b)\right)\right) \right]$

$\log(1 + e^{-y_i (w^T x_i + b)})$

$\Rightarrow \log\left(1 + e^{-\hat{y}_i \cdot y_i}\right)$

$-\log\left(\frac{1}{1 + e^{-\hat{y}_i \cdot y_i}}\right)$

$-\log\left(\sigma(\hat{y}_i \cdot y_i)\right)$