Lec 12: Support Vector Machine (SVM)

Ailin Zhang

2022-10-24

Agenda

- Classification, outcome, and logistic loss
- Perceptron and margin
- SVM
- Primal form
- Dual form

Warm up: Logsitic Regression

obs	$\mathbf{X}_{n\times p}$	$\mathbf{Y}_{n\times 1}$
1		
2		
...		
i	$X_i^\top$	y_i
...		
n		

Let $\eta_i = X_i^\top \beta$ be the score, then

$$p_i = \sigma(\eta_i) = \frac{1}{1+e^{-\eta_i}} = \frac{1}{1+e^{-X_i^\top \beta}} = \frac{e^{X_i^\top \beta}}{1+e^{X_i^\top \beta}},$$

$$1 - p_i = \frac{1}{1+e^{X_i^\top \beta}}$$

$\eta_i = \log(\frac{p_i}{1-p_i}) = logit(p_i) =$ log odds ratio

IRLS

$$l(\beta) = -\sum_{i=1}^{n} \hat{w}_i (\hat{y}_i - x_i^T \Delta\beta)^2.$$

Where $w_i = p_i(1-p_i)$, $\hat{y}_i = \frac{\hat{e}_i}{\hat{w}_i}$

Recall linear regression:

$$\sum_{i=1}^{n} (y_i - x_i^T \beta)^2.$$

$$\begin{aligned}
\beta^{(t+1)} &= \beta_t + (\sum_{i=1}^{n} w_i X_i X_i^\top)^{-1} (\sum_{i=1}^{n} w_i X_i \hat{y}_i) \\
&= \left(\sum_{i=1}^{n} w_i X_i X_i^\top\right)^{-1} \left[\sum_{i=1}^{n} w_i X_i \left(X_i^\top \beta^{(t)} + \frac{y_i - p_i}{w_i}\right)\right].
\end{aligned}$$

Perceptron

- Perceptron is a binary classifier: $\hat{y}_i = sign(\mathbf{X}_i^\top \beta)$.
- Logistic Regression is a soft version of perception
- Logistic Regression is also a generalized linear model (GLM)

Perceptron Model

The perceptron model $y_i = \mathrm{sign}(X_i^\top \beta)$, where $y_i \in \{+1, -1\}$

The gradient learning algorithm can be modified into the perceptron algorithm:

Starting from $\beta_0 = 0$,

$$\beta^{(t+1)} = \beta^{(t)} + \sum_{i=1}^{n} \delta_i y_i X_i,$$

where $\delta_i = 1(y_i \neq \mathrm{sign}(X_i^\top \beta^{(t)}))$ to determine whether $\beta^{(t)}$ makes a mistake in classifying y_i.

The algorithm can be considered the gradient descent algorithm for the loss function

$$loss(\beta) = \sum_{i=1}^{n} \max(0, -y_i X_i^\top \beta)$$

Margin

$$loss(\beta) = \sum_{i=1}^{n} \max(0, -y_i X_i^\top \beta) = \sum_{i=1}^{n} \max(0, -margin_i),$$

- $\max(0, -margin_i) = 0$ if $margin_i \geq 0$, i.e., no mistake is made
- $\max(0, -margin_i) = -margin_i$ if $margin_i < 0$.

Again the algorithm learns from the mistakes.

Support Vector Machine - Motivation

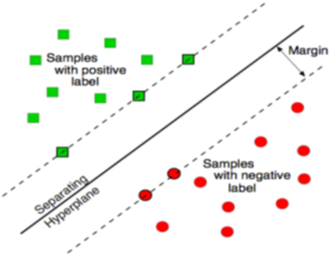
Consider the perceptron $y_i = \mathrm{sign}(X_i^\top \beta)$:

1. It separates the positive examples and negative examples by projecting the data on vector β,
2. It separates the examples by a hyperplane that is perpendicular to β.

If the positive examples and negative examples are separable, there are be many separating hyperplanes.

We want to choose the one with the **maximum margin** in order to guard against the random fluctuations in the unseen testing examples.

Support Vector Machine



The idea of support vector machine (SVM) is to find the β, so that

1. for positive examples $y_i = +$, $X_i^\top \beta \geq 1$,
2. for negative examples $y_i = -$, $X_i^\top \beta \leq -1$.

Here we use $+1$ and -1, because we can always scale β.

The decision boundary is decided by the training examples that lies on the margin. Those are the support vectors.

Support Vector Machine

Let u be an unit vector that has the same direction as β. $u = \dfrac{\beta}{|\beta|}$.

Suppose X_i is an example on the margin (i.e., support vector), the projection of X_i on u is

$$\langle X_i, u \rangle = \langle X_i, \frac{\beta}{|\beta|} \rangle = \frac{X_i^\top \beta}{|\beta|} = \frac{\pm 1}{|\beta|}.$$

So the margin is $1/|\beta|$. In order to maximize the margin, we should minimize $|\beta|$ or $|\beta|^2$. Hence, the SVM can be formulated as an optimization problem as follows:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}|\beta|^2, \\ \text{subject to} \quad & y_i X_i^\top \beta \geq 1, \forall i. \end{aligned}$$

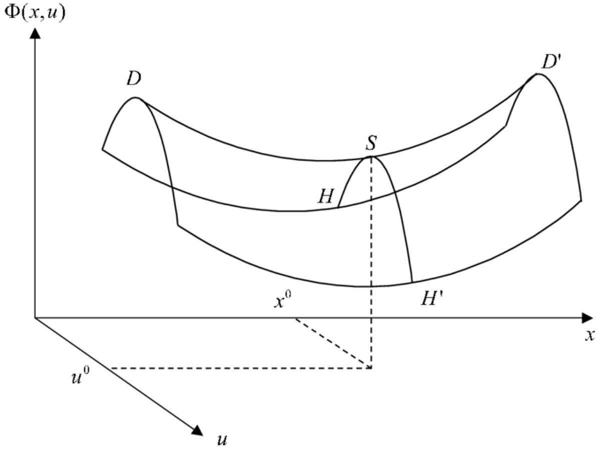
Recall $X_i^\top \beta$ is the score, and $y_i X_i^\top \beta$ is the individual margin of observation i. This is the **primal form** of SVM.

Dual Form: Lagrange Multiplier

Let $\alpha = (\alpha_1, \alpha 2, \cdots, \alpha_n)$, where $\alpha_i \geq 0$
$L(\beta, \alpha) = \frac{1}{2}|\beta|^2 + \sum_{i=1}^{n} \alpha_i (1 - y_i X_i^\top \beta)$

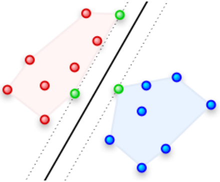
The idea is to solve an unconstrained problem because it is easier to solve.

Dual Form: Lagrange Multiplier and saddle point



Dual Form

The primal form of SVM is max margin, and the dual form of SVM is min distance.



max margin = min distance

The margin between the two sets is defined by the minimum distance between two.

Dual Form - Convex Hull

Let $X_+ = \sum_{i\in+} c_i X_i$ and $X_- = \sum_{i\in-} c_i X_i$ ($c_i \geq 0, \sum_{i\in+} c_i = 1, \sum_{i\in-} c_i = 1$) be two points in the positive and negative convex hulls. The margin is $\min \left|X_+ - X_-\right|^2$.

$$
\begin{aligned}
\left|X_+ - X_-\right|^2 &= \left|\sum_{i\in+} c_i X_i - \sum_{i\in-} c_i X_i\right|^2 \\
&= \left|\sum_i y_i c_i X_i\right|^2 \\
&= \sum_{i,j} c_i c_j y_i y_j \langle X_i, X_j\rangle, \\
\text{subject to} \quad & c_i \geq 0, \sum_{i\in+} c_i = 1, \sum_{i\in-} c_i = 1.
\end{aligned}
$$

Solvable with sequential minimal optimization