ch09_notes

August 10, 2019

Table of Contents

---

1 Support Vector Machines

1.1 Maximal Margin Classifier

1.1.1 What Is a Hyperplane?

- A ***hyperplane*** in $\mathbb{R}^p$ is an affine subspace of dimension $p-1$. Every hyperplane is the set of solutions X to $\beta^\top X = 0$ for some $\beta \in \mathbb{R}^p$.
- A hyperplane $\beta^\top X = 0$ partitions $\mathbb{R}^p$ into two halfspaces:

$$H_+ = \{X \in \mathbb{R}^p \mid \beta^\top X > 0\}$$
$$H_- = \{X \in \mathbb{R}^p \mid \beta^\top X > 0\}$$

corresponding to either side of the plane, or equivalently,

$$H_+ = \{X \in \mathbb{R}^p \mid \text{sgn}(\beta^\top X) = 1\}$$
$$H_- = \{X \in \mathbb{R}^p \mid \text{sgn}(\beta^\top X) = -1\}$$

1.1.2 Classification Using a Separating Hyperplane

- Given data (x_i, y_i), $i = 1, \ldots n$ with response classes $y_i \in \{\pm 1\}$, a hyperplane $\beta^\top X = 0$ is ***separating*** if

$$\text{sgn}(\beta^\top x_i) = y_i$$

for all i. - Given a separating hyperplane, we may predict

$$\hat{y}_i = \text{sgn}(\beta^\top x_i)$$

1.1.3 The Maximal Margin Classifier

- Separating hyperplanes are not unique (if one exists then uncountably many exist). A natural choice is the ***maximal margin hyperplane*** (or ***optimal separating hyperplane***)
- The ***margin*** is the minimal perpendicular distance to the hyperplane over the sample points

  $$M = \min_i \{\ ||x_i - Px_i||\ \}$$

  where P is the projection matrix onto the hyperplane.
- The points (x_i, y_i) "on the margin" (where $||x_i - Px_i|| = M$) are called ***support vectors***

1.1.4 Construction of the Maximal Margin Classifier

The maximal margin classifier is the solution to the optimization problem:

$$\begin{aligned} &\underset{\beta}{\text{argmax}}\ M \\ &\text{subject to } ||\,\boldsymbol{\beta}\,|| = 1 \\ &\qquad\qquad \mathbf{y}^\top (X\boldsymbol{\beta}) \geqslant \mathbf{M} \end{aligned}$$

where $\mathbf{M} = (M, \ldots, M) \in \mathbb{R}^n$ 78

1.1.5 The Non-separable Case

- The maximal margin classifier is a natural classifier, but a separating hyperplane is not guaranteed to exist
- If a separating hyperplane doesn't exist, we can choose an "almost" separating hyperplane by using a "soft" margin.

1.2 Support Vector Classifiers

1.2.1 Overview of the Support Vector Classifier

- Separating hyperplanes don't always exist, and even if they do, they may be undesirable.
- The distance to the hyperplane can be thought of as a measure of confidence in the classification. For very small margins, the separating hyperplane is very sensitive to individual observations – we have low confidence in the classification of nearby observations.
- In these situations, we may prefer a hyperplane that doesn't perfectly separate in the interest of:
  - Greater robustness to individual observations
  - Better classification of most of the training observations
- This is achieved by the ***support vector classifier*** or ***soft margin classifier*** 79

1.2.2 Details of the Support Vector Classifier

- The support vector classifier is the solution to the optimization problem:

$$
\begin{aligned}
&\underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, M \\
\text{subject to } &\|\boldsymbol{\beta}\| = 1 \\
&y_i(\boldsymbol{\beta}^\top x_i) \geqslant M(1-\epsilon_i) \\
&\epsilon_i \geqslant 0 \\
&\sum_i \epsilon_i \leqslant C
\end{aligned}
$$

where $C \geqslant 0$ is a tuning parameter, M is the margin, and the ϵ_i are slack variables. 80

- Observations on the margin or on the wrong side of the margin are called ***support vectors***

1.3 Support Vector Machines

1.3.1 Classification with Non-Linear Decision Boundaries

- The support vector classifier is a natural choice for two response classes when the class boundary is linear, but may perform poorly when the boundary is non-linear.
- Non-linear transformations of the features will lead to a non-linear class boundary, but enlarging the feature space too much can lead to intractable computations.
- The support vector machine enlarges the feature space in a way which is computationally efficient.