

サポートベクター回帰についてのメモ

岡田 大 (Okada Masaru)

November 8, 2025

Abstract

機械学習の tips として、回帰手法の一つのサポートベクター回帰について解説します。

1 線形サポートベクター回帰

まずはシンプルな線形サポートベクター回帰から始めます。

線形サポートベクターでは線形モデルのみを扱います。

この範囲では線形予測モデルは次のように表現されます。

$$h_{\theta}(x) = \theta^T x + \theta_0$$

ここで、

$$x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$$

は特徴量（訓練データ）、

$$\theta = (\theta_0, \theta_1, \theta_2, \dots, \theta_n)^T \in \mathbb{R}^{n+1}$$

はバイアスパラメータと呼ばれるモデルの係数でそれぞれ n 次元、 $n+1$ 次元のベクトルです。

サポートベクター回帰による解（モデル、すなわちバイアスパラメータ θ ）は、次で表される損失関数 $J(\theta)$ を（条件付きで）最小化する問題の解として表されます。

$$J(\theta) = C \sum_{i=1}^n \left(\xi^{(i)} + \hat{\xi}^{(i)} \right) + \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

この第1項はマージン違反の損失関数、第2項はL2正則化の損失関数と言います。ここで $\xi, \hat{\xi} \in \mathbb{R}^n$ となるベクトルです。

ただし、損失関数 $J(\theta)$ 最小化問題の制約条件としてサポートベクター回帰では以下の条件が付随します。

$$y^{(i)} \leq h_{\theta}(x^{(i)}) + \varepsilon + \xi^{(i)}$$

$$y^{(i)} \geq h_{\theta}(x^{(i)}) - \varepsilon - \hat{\xi}^{(i)}$$

$$\xi^{(i)} \geq 0$$

$$\hat{\xi}^{(i)} \geq 0$$

ここで $y \in \mathbb{R}^n$ は正解データです。

C と ε は人間が手でチューニングする1次元の値のパラメータになります。（こういった機械学習モデルに付随するパラメータをハイパープラメータと呼びます。）

すなわち、（単に式を移項しただけですが、）

$$\xi = y - (h_{\theta}(x) + \varepsilon)$$

$$\hat{\xi} = (h_{\theta}(x) - \varepsilon) - y$$

はそれぞれ正解データと予測モデルの上側または下側の差分であり、そこに ε の幅（これを「チューブ」と言います）を持たせた領域を最小化問題を解くことで予測します。

1.1 ハイパーパラメータ C

サポートベクター回帰のアルゴリズムの誤差は ε チューブの上側と下側でそれぞれ $\xi, \hat{\xi}$ だけ発生しますが、例えばチューブ上部の誤差 ξ は正解 y と予測モデル $h_\theta(x) + \varepsilon$ の残差になっています。 C の値が大きいほど正解 y がチューブから離れたときの残差 ξ が大きくなり、マージン違反の損失関数の影響が大きくなります。

1.2 ハイパーパラメータ ε

$J(\theta)$ の条件付き最小化問題はラグランジュの未定乗数法を用いることが有効です。ラグランジュの未定乗数法を用いると、ラグランジアンと呼ばれる関数 L が次のように書かれることができます。

$$L(a, \hat{a}) = -\frac{1}{2} \sum_{i,j} (a^{(i)} - \hat{a}^{(i)})(a^{(j)} - \hat{a}^{(j)})K(x^{(i)}, x^{(j)}) - \varepsilon \sum_i (a^{(i)} + \hat{a}^{(i)}) + \varepsilon \sum_i (a^{(i)} - \hat{a}^{(i)})y^{(i)}$$

ここで K はカーネル関数と呼ばれる関数です。線形サポートベクターでは

$$K(x^{(i)}, x) = ((x^{(i)})^T x)$$

のようにベクトルの内積になります。 $a, \hat{a} \in \mathbb{R}^n$ は ε チューブの外側（それぞれチューブの上側、下側）の訓練データで、サポートベクターと呼ばれます。このラグランジアン $L(a, \hat{a})$ の最小値を求める上でモデルが得られるわけですが、ここから ε チューブの外側の訓練データ（サポートベクター）のみでモデルが生成されることが分かります。

未定乗数法から $J(\theta)$ を極小化する θ は以下の式で表されます。

$$\theta = \sum_i (a^{(i)} - \hat{a}^{(i)})x^{(i)}$$

これを予測モデル

$$h_\theta(x) = \theta^T x + \theta_0$$

に代入すると、

$$h_\theta(x) = \sum_i (a^{(i)} - \hat{a}^{(i)})((x^{(i)})^T x) + \theta_0$$

が得られます。

2 サポートベクター回帰

これまで見てきたカーネル関数が内積の場合は線形サポートベクター回帰と呼びます。ここからはサポートベクター回帰を非線形な場合にも拡張することを考えます。先ほどの予測モデルの式に出てきた内積を次のようにカーネル関数 K を用いて書き換えます。

$$h_\theta(x) = \sum_i (a^{(i)} - \hat{a}^{(i)})K(x^{(i)}, x) + \theta_0$$

カーネル関数の具体例としては

線形カーネル：

$$K(x^{(i)}, x) = ((x^{(i)})^T x)$$

ガウシアンカーネル：

$$K(x^{(i)}, x) = \exp(-\gamma|x^{(i)} - x|^2) \quad \text{where } \gamma = \frac{1}{2\sigma^2}$$

ここで σ^2 はガウシアンの分散です。

他にも多項式カーネルやシグモイドカーネルも利用されます。

カーネルの関数形もハイパーパラメータの一つです。

3 時系列データの推移予測のハイパーパラメータ

以上がざっくりとしたサポートベクター回帰の説明になります。

例として今回用いたハイパーパラメータは以下のようになっています。

Table 1: 時系列データの推移予測のハイパーパラメータ

C	ε	K	γ
100	7	ガウシアンカーネル	$(\dim x)^{-1}$

パラメータ空間の中を走査して誤差が最も少ないパラメータセットを選ぶというのが通常やられることだと思いますが、今回は決め打ちでやりました。

本気でやるなら曜日ごと、週ごとに回帰を用いたり、土日祝だけフラグを立てる等でさらに予測精度が改善されると思いますが、今回はあえてここまでしか詰めていません。

これ以上まじめに精度を出すにはそもそもサポートベクター回帰とは異なる手法の方が向いていそうと感じた為です。

次には XGBoost、LightGBM といった別の方法で試してみて、また今回のように簡単に解説する記事を書きたいと思います。