

3.3

y_i ($i: 1, \dots, T$) : ビールの売り上げ本数 ← 目標値
 $\begin{cases} x_{i1} & (1, 2, \dots, 30) & : & \text{日付} \\ x_{i2} & & : & \text{気温} \\ x_{i3} & & : & \text{湿度} \end{cases}$ ← 特徴量
 $x_i^T = (x_{i1}, x_{i2}, x_{i3})$

(1) サンプル

(2)
$$f(x_i) = w_0 + \sum_{m=1}^4 \sum_{d=1}^3 w_{md} x_{id}$$

(3) 訓練誤差

$$\text{Training Err} = \frac{1}{|N|} \sum_{i=1}^N (y_i - f(x_i))^2$$

(4) 過学習

- 日付に周期性があること
- 天気の影響がある可能性
- たとえば 目標値の分散が大きくなり、当てずっぽう

(5) 過学習のせいで、正しい予測をすることができなくなった

- 気温が急激に下がった → 特徴量にノイズが入ったため ← 特徴量ではなかった
- 給料日の前日であった → 気温が特徴量外に存在したため ← イベントを表すフラグを特徴量に入れる

(6) k- fold 交差検証を用いてテスト誤差を評価し、テスト誤差が最小となる多項式回帰のモデルを学習させる。また、リッジ回帰により過度な複雑さを抑え、多項式回帰のモデルを考える。

(7) イベント (祝日、給料日) の日とそうでない日を区別し、イベント時に売れる本数外

多々あるようなモデルを設計する

月、日曜日を区別

- テレビでビールを飲んでいるシーンがあった
- Xメディアでビールが取り上げられた場合 伸びる → 過去7日中のビール消費
- ニュース数と特徴量と
- シーンの放映数