数理工学のすすめ
第 3,4 回「機械学習とその数理」レポート課題

2017-4-27

鈴木大慈
e-mail: taiji@mist.i.u-tokyo.ac.jp

以下の課題のうち，いずれか一つを解け．

課題１

1. 平均 $\mu \in \mathbb{R}$，分散 $\sigma^2 > 0$ の正規分布は次の確率密度関数を持つ分布と定義される:

$$p(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

   これが実際に平均 μ，分散 σ^2 を持つことを示せ．ただし，$p(x;\mu,\sigma^2)$ が確率密度関数であること ($\int_{-\infty}^{\infty} p(x;\mu,\sigma^2)\mathrm{d}x = 1$) は断りなしに用いてよい．

2. $Y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times p}$, $\epsilon \in \mathbb{R}^n$, $\beta^* \in \mathbb{R}^p$ として，線形回帰モデル

$$Y = X\beta^* + \epsilon$$

   を考える．ここで，$(\epsilon_i)_{i=1}^n$ は互いに独立で，それぞれ平均 0，分散 σ^2 の確率変数であるとする．β^* の推定量 $\hat{\beta}$ に対して，

$$\mathrm{E}_\epsilon[\|X\hat{\beta} - X\beta^*\|^2] = \mathrm{E}_\epsilon[\|X\hat{\beta} - X\mathrm{E}_\epsilon[\hat{\beta}]\|^2] + \|X\mathrm{E}_\epsilon[\hat{\beta}] - X\beta^*\|^2$$

   なるバイアス-バリアンス分解が可能であることを示し，リッジ回帰推定量に対してバイアス項とバリアンス項を評価せよ．特に，正則化パラメータ λ が 0 または ∞ の時にそれぞれどのような値になるか述べよ．

課題２ 講義で扱ったリッジ回帰や他の機械学習手法を実際にデータに適用し，過学習の現象を観察せよ．なお，データは人工データを用いても実データを用いても良い．
(参考：scikit-learn の例題 “Underfitting vs. Overfitting” `https://goo.gl/2GbrHg`)

課題３ 機械学習および統計的データ解析の現実的な応用例について文献を調べてまとめよ．また，その応用例ではどのような数学が有用であるか考えを述べよ．
ただし，参考とした資料は明記すること．