

# 数理・人工知能のフロンティアと社会価値創造 課題 2 日目

理学部 4 年 1SC22317Y 照屋佑喜仁

2025 年 6 月 6 日

レポート課題の例の場合を考えてみる。つまり

- 糖尿病患者などの血糖値測定では、食後の血糖値がある閾値を超えるかどうかを予測したい。
- データは大半が食後ではない低血糖の状態のため、血糖値が低い状態の精度が高い一方、重要な高血糖の状態の精度が低い問題があります。高血糖の方が精度が高いモデルを作りたい。

というような状況でどのような loss function を設定するか考える。

このような状況で、loss function としてよく使われるような関数 (例えば平均 2 乗誤差 MSE など) を用いてしまうと、異常値に反応しないモデルになる可能性がある。そこで私は loss function に正則化を導入しようと考えた。非線形の回帰分析等では、データに当てはまりが良すぎると過学習してしまい予測に使えないため正則化 (罰則を設ける) するということをゼミで学んだ。それを使うことができないだろうかという目論見である。

以下を提案する。簡単のため学習データは (時間で) 分割したものを考える。

$$\text{Loss}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^n \mathbf{1}_{y_i \geq T} \mathbf{1}_{y_i \geq \hat{y}_i}$$
$$\mathbf{1}_{y_i \geq T} = \begin{cases} 1 & y_i \geq T \\ 0 & y_i < T \end{cases}$$
$$\mathbf{1}_{y_i \geq \hat{y}_i} = \begin{cases} 1 & y_i \geq \hat{y}_i \\ 0 & y_i < \hat{y}_i \end{cases}$$

ただし  $n$  をデータ数とし、 $\mathbf{y}, \hat{\mathbf{y}}$  を実際のデータと予測データ、 $T$  を腎閾値とし、 $\lambda$  は罰則の重みパラメータとする。

これは通常のアVERAGE 2 乗誤差に罰則を加えたもので、腎閾値を超えていたときに実際の値より下回って予測していた場合に罰則を設けている。ハイパーパラメータをうまく調整することで、低血糖時も高血糖時 (閾値を超える場合のみに限ってしまうが) も予測できるのではと考えた。