

EEE950 - Redes Neurais Artificiais: Teoria e Aplicações

Exercício 2 - Regularização com redes ELM

José Geraldo Fernandes
Escola de Engenharia
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil

I. INTRODUÇÃO

Implementou-se uma rede ELM, com função de ativação tangente hiperbólica, para estudar efeitos de *overfitting* e regularização.

Dada uma projeção aleatória Z de tamanho p , resolve-se o problema de mínimos quadrados como na Equação 1.

$$\begin{aligned} H &= \tanh(XZ) \\ W &= (H^T H + \lambda I_p)^{-1} H^T y \\ \hat{y} &= \tanh(HW) \end{aligned} \quad (1)$$

O tamanho da projeção aleatória p controla a complexidade do modelo, enquanto a regularização é em função do parâmetro λ . O balanço desses parâmetros é necessário para solução do problema.

II. XOR

Utilizou-se o problema sintético da disjunção exclusiva (XOR) para explorar o *trade-off* de p e λ .

Um primeiro experimento com p baixo foi suficiente para modelar a função sem regularização, $\lambda = 0$, como na Figura 1.

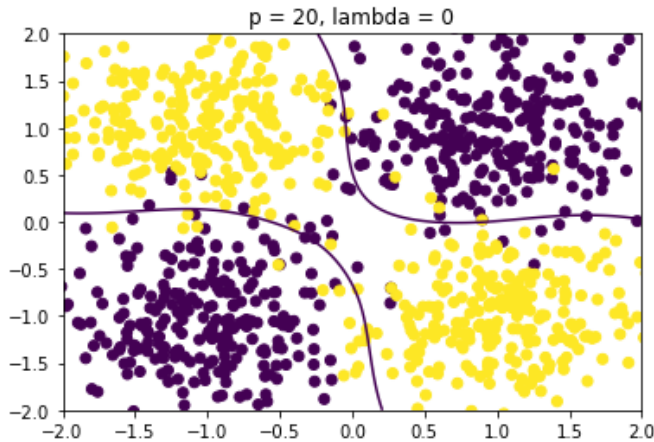


Figura 1. Superfície de separação no problema XOR, ajustada por uma ELM com 20 neurônios escondidos e sem regularização.

Aumentando a complexidade do modelo, parâmetro p , observa-se o efeito de *overfitting* na superfície do discriminador, como na Figura 2. Esse fenômeno é caracterizado pela sinuosidade da curva.

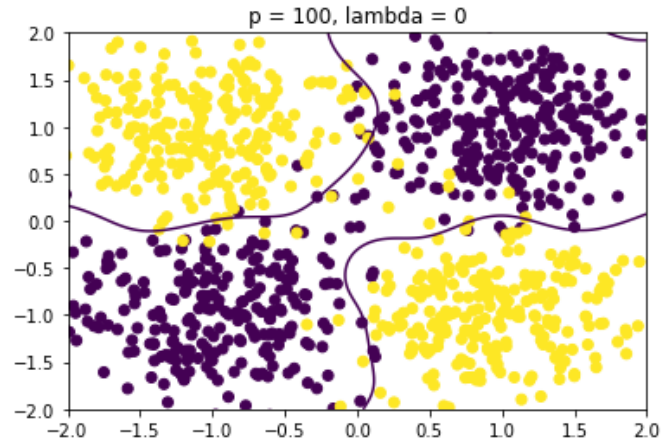


Figura 2. Superfície de separação no problema XOR, ajustada por uma ELM com 100 neurônios escondidos e sem regularização.

Em contrapartida, injetando regularização nesse modelo, parâmetro λ , é possível modelar adequadamente a função mesmo com complexidade alta, como mostra a Figura 3.

Como é comumente difícil determinar um p ótimo que ajuste a função adequadamente, a técnica de regularização é uma proposta poderosa para esse problema.

III. STATLOG GERMAN CREDIT

Utilizou-se, agora, um problema de classificação binário real para verificar se o comportamento observado é semelhante. Para isso, os dados utilizados são o *Statlog German Credit*.

Como a visualização da curva de separação é inacessível em alta dimensão, demonstra-se os efeitos de *overfitting* e regularização a partir do custo do modelo em um conjunto de treinamento e outro de teste. Para tal, aplicou-se validação cruzada em $k = 10$ folds, e o custo quadrático médio (MSE).

A Figura 4 mostra um experimento sem regularização e aumentando p exponencialmente. Como esperado, a medida que

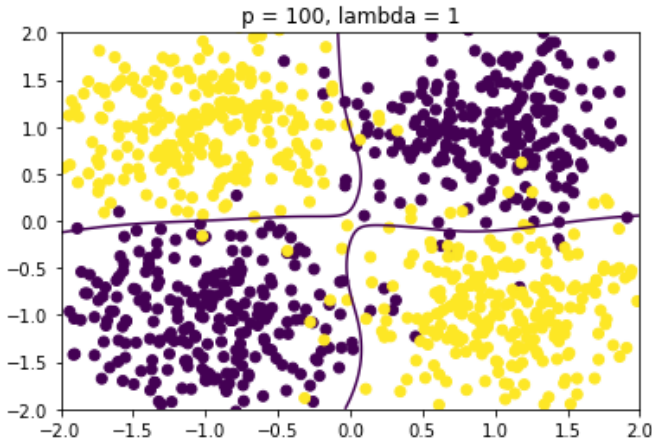


Figura 3. Superfície de separação no problema XOR, ajustada por uma ELM com 100 neurônios escondidos mas regularizado.

a complexidade aumenta o custo do conjunto de treinamento cai, contudo, a partir de um valor ótimo o custo no conjunto de teste diverge, caracterizando *overfitting*.

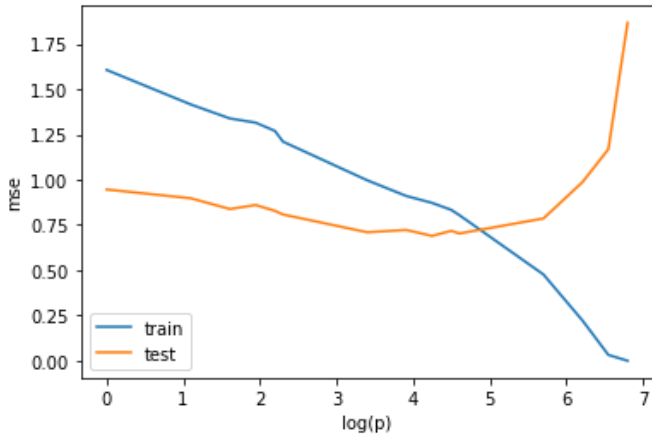


Figura 4. Curvas do custo de treinamento e teste variando a complexidade do modelo.

Selecionando um desses modelos inadequados, $p = 300$, variou-se o parâmetro λ exponencialmente, como na Figura 5.

Note que o ajuste de λ faz convergir as curvas de custo, modelando, portanto, a função adequadamente mesmo com p evidencialmente ruim, algo comum na prática.

IV. CONCLUSÃO

Mostrou-se a capacidade da rede ELM de modelar função de separação não-lineares, como o problema sintético de disjunção exclusiva e um *dataset* real. No entanto, esse modelo pode facilmente cair em *overfitting* dada situações práticas comuns.

Uma maneira eficiente de tratar esse problema é a técnica de regularização. Foi demonstrado que mesmo modelos que

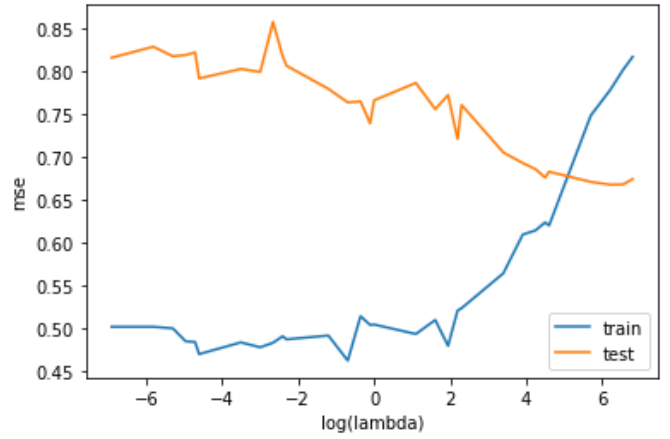


Figura 5. Curvas do custo de treinamento e teste variando a regularização.

mostraram *overfitting* sem regularização podem ser modificados para modelar a função adequadamente.