

Artigo 01 - RNAs PPGEE - Segundo semestre de 2022

23 de setembro de 2022

1 Caracterização do problema

É sabido que a capacidade efetiva de uma rede neural depende das restrições impostas ao seu espaço de soluções, podendo estas serem determinadas pelo número de parâmetros do modelo ou por outras formas de restrições à busca neste espaço. O trabalho de Geman e colaboradores [GBD92] mostra claramente como a variabilidade da família de modelos aproximadores influencia na qualidade da aproximação. A decomposição da esperança do erro quadrático, apresentada no trabalho de Geman, resulta na Equação 1, que é conhecida por apresentar a decomposição do erro de aproximação $E[(f(\mathbf{x}, \mathbf{w}) - E[y|x])^2]$ em dois termos, que representam o viés e a variância da família de modelos. O objetivo da aproximação é aproximar o valor esperado de saída, $E[y|x]$, assim, conseqüentemente, ambos os termos $(E[f(\mathbf{x}, \mathbf{w})] - E[y|x])^2$ (viés) e $E[(f(\mathbf{x}, \mathbf{w}) - E[f(\mathbf{x}, \mathbf{w})])^2]$ (variância) devem ser minimizados.

$$E[(f(\mathbf{x}, \mathbf{w}) - E[y|x])^2] = \underbrace{(E[f(\mathbf{x}, \mathbf{w})] - E[y|x])^2}_{\mu_{\{f(\mathbf{x}, \mathbf{w})\}}} + \underbrace{E[(f(\mathbf{x}, \mathbf{w}) - E[f(\mathbf{x}, \mathbf{w})])^2]}_{\sigma_{\{f(\mathbf{x}, \mathbf{w})\}}^2} \quad (1)$$

Métodos de regularização [Wah94, TA77], por exemplo, restringem o espaço de busca a soluções de menor magnitude de pesos, dependendo do valor de λ , conforme Equação 2.

$$J = \sum_{i=1}^N (y_i - f(\mathbf{x}_i, \mathbf{w}))^2 + \lambda \sum_{j=1}^p w_j^2 \quad (2)$$

em que N é o número de amostras e p o número total de parâmetros do modelo, ou pesos, caso o modelo seja uma rede neural.

A Figura 1 mostra o resultado de um experimento simples em que a variância dos valores de \mathbf{w} decresce em função do aumento à restrição aos seus valores no

espaço de soluções. Para a realização do experimento foi realizada a amostragem de 1000 valores de w de uma distribuição normal com média 5 e desvio padrão 1. A variância desta amostra foi estimada e armazenada para posterior comparação com os valores que serão estimados após a restrição no espaço. Como existe um modelo neural para cada valor de w neste exemplo hipotético, a variância da amostra é uma medida da variância dos modelos, segundo termo da Equação 1. Em seguida foi imposta uma restrição de valores de w em torno da média da distribuição, restrição esta que variou de 0 a 4.9 e, para cada valor da restrição foi calculada a variância dos valores de w da amostra que estavam dentro dos limites da restrição. Os resultados apresentados no gráfico são a razão entre a variância estimada no espaço com restrição e aquela obtida inicialmente, sem restrição de magnitude. O gráfico da Figura 1 mostra, assim, como a variância dos valores de w , ou dos modelos correspondentes, se comporta à medida em que a restrição aumenta. A restrição de norma, ou de magnitude, é o princípio básico dos métodos baseados em regularização [TA77] e também daqueles baseados em aprendizado multi-objetivo [TBTR00].

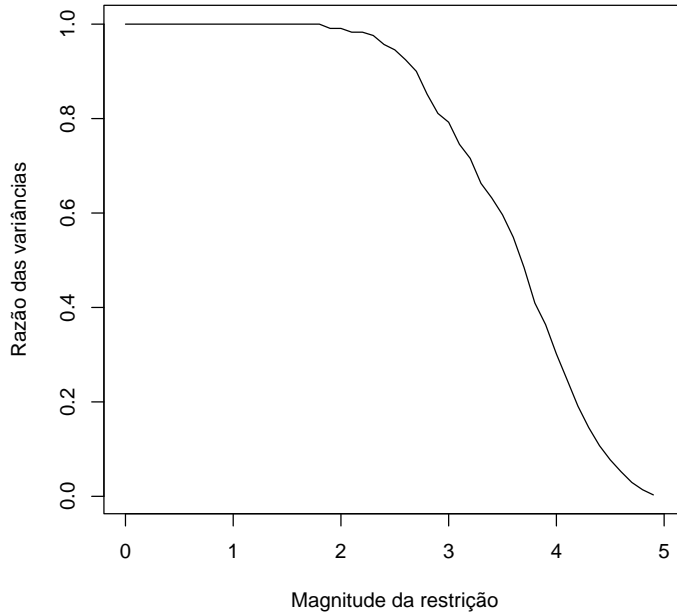


Figura 1: Comportamento da variância do vetor de pesos em função da restrição de magnitude imposta ao espaço de soluções.

1.1 Regularização de redes ELM e RBF

O material indicado aos alunos, incluindo Notas de Aulas, exercícios, etc, descreve formulações para regularização de ELMs e RBFs, assim como para a adição e remoção de neurônios e amostras de treinamento. Neste trabalho o aluno deverá considerar este material como base para a realização de experimentos e propostas de novos métodos e estratégias de treinamento destes dois modelos de redes neurais. O objetivo é o desenvolvimento de métodos de aprendizado com controle de capacidade utilizando regularização e/ou estratégias de adição/remoção de amostras e neurônios. Estratégias possíveis são, por exemplo:

- Métodos construtivos[FL88, Ree93, Kar90] visando à adição/remoção de neurônios onde o foco pode ser em "quais neurônios adicionar ou retirar".
- Treinamento com ruído, e.g. adicionar amostras sintéticas na região de separação, visando à suavização da resposta do modelo[Bis95].
- Estratégias para a seleção do fator λ da regularização.
- Outros, tema livre proposto e discutido com os professores em sala de aula.

2 O que deve ser entregue

- **Formato:** Formato de artigo (Ref: padrão IEEE), duas colunas, máximo de 8 páginas.
- **Objetivo/tema:** Conforme seção anterior.
- **Estrutura:** O texto do trabalho deve conter pelo menos Título, Introdução, Revisão de literatura, Metodologia, Resultados, Discussões, Conclusões e Bibliografia.
- **Entrega:** via Moodle na data indicada.

Referências

- [Bis95] Chris M Bishop. Training with noise is equivalent to tikhonov regularization. *Neural computation*, 7(1):108–116, 1995.
- [FL88] S. E. Fahlman and C. Lebiere. The cascade-correlation learning architecture. In R. P. Lippmann, J. E. Moody, and D. S. Touretzky, editors, *Advances in Neural Information Processing Systems 2*. Morgan Kaufmann, 1988.
- [GBD92] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias-variance dilemma. *Neural Computation*, 4:1–58, 1992.

- [Kar90] E. D. Karnin. A simple procedure for pruning back-propagation trained neural networks. *IEEE Transactions on Neural Networks*, 1(2):239–242, 1990.
- [Ree93] R. Reed. Pruning algorithms — A survey. *IEEE Transactions on Neural Networks*, 4(5):740–746, 1993.
- [TA77] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. V.H. Winston & Sons, John Wiley & Sons, Washington D.C., 1977. Translation editor Fritz John.
- [TBTR00] R.A. Teixeira, A.P. Braga, R.H.C. Takahashi, and R. Rezende. Improving generalization of mlps with multi-objective optimization. *Neurocomputing*, 35:189–194, 2000.
- [Wah94] Grace Wahba. Generalization and regularization in nonlinear learning systems. Technical Report 921, University of Wisconsin, May 1994.