

Comparação de Modelos de Representação Latente

José Geraldo Fernandes
Escola de Engenharia
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil
josegeraldof@ufmg.br

I. INTRODUÇÃO

Aprendizado de representação é uma propriedade muito importante que justifica alguma parte do grande sucesso de modelos profundos [1]. Em uma rede neural a última camada é apenas um discriminador linear, cabe, portanto, às camadas anteriores aprender uma representação dos dados que modele o problema de classificação linearmente. Em redes convolucionais, há uma semelhança em o quão fundo está a camada e a complexidade de *features* que o mesmo captura.

Estudar esse espaço latente, portanto, é de grande utilidade para a aplicação desses modelos e uma vez com esse entendimento consolidado uma direção pode ser induzida além da minimização do erro. Informação de vizinhança já foi explorado [2] como informação *a priori* para esse direcionamento e outros aspectos como esparsividade e sobreposição podem ser úteis.

Esse trabalho estuda a representação aprendida de uma rede neural a partir de uma avaliação de sobreposição.

II. REVISÃO BIBLIOGRÁFICA

A. Grafo de Gabriel

O Grafo de Gabriel [3] é uma construção a partir de um conjunto de pontos \mathcal{S} , que define os vértices do grafo, para outro conjunto de arestas \mathcal{E} tal que dois pontos $\mathbf{x}_i, \mathbf{x}_j$ são adjacentes, definem uma aresta, se não há outro ponto de \mathcal{S} dentro da hipersfera definida com a distância entre os dois pontos $\mathbf{x}_i, \mathbf{x}_j$ como diâmetro, como na Equação 1.

$$\begin{aligned} (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{E} \Leftrightarrow \\ \delta^2(\mathbf{x}_i, \mathbf{x}_j) \leq [\delta^2(\mathbf{x}_i, \mathbf{x}_k) + \delta^2(\mathbf{x}_j, \mathbf{x}_k)] \quad \forall \mathbf{x}_k \in \mathcal{S} \end{aligned} \quad (1)$$

Onde δ é a métrica de distância na construção e comumente definida como a distância euclidiana.

A construção desse grafo é uma técnica da Geometria Computacional e emprestada para os problemas de aprendizado como uma forma de expressar as características de vizinhança da estrutura do conjunto de dados.

1) *Métrica de Qualidade*: Baseado na diferença de classes entre o vértice e as amostras conectadas por suas arestas, define-se uma grandeza Q que representa a qualidade dessa amostra como na Equação 2, onde V representa o número de arestas e o somatório itera sobre as amostras vizinhas, calculando o número de vizinhos dissidentes, com rótulo diferente, ponderado pelo inverso de sua distância.

$$Q(\mathbf{x}_i) = 1 - \frac{\sum_j |y_i - y_j| \exp(-\delta(\mathbf{x}_i, \mathbf{x}_j))}{V(\mathbf{x}_i)} \quad (2)$$

Amostras na região de fronteira e sobreposição terão sua qualidade afetada quão maior for a mistura. Também, amostras com qualidade perfeita $Q(\mathbf{x}_i) = 1$ estão em regiões predominantes e longe da fronteira.

Neste trabalho, contudo, é interesse utilizar esse indicador como uma forma de avaliar o espaço latente aprendido por uma rede neural.

III. METODOLOGIA

A. Base de Dados

Para avaliação dos índices de qualidade dos agrupamentos seleciona-se um conjunto de 15 *datasets* padrão do repositório UCI [4] em todos os testes. Esses são separados em 10 partições determinadas para validação cruzada *k-fold*.

Para confiabilidade do resultado, as bases de dados selecionadas são amplamente utilizadas em aplicações semelhantes. Como pré-processamento, fez-se uma normalização de todos os atributos e conversão dos categóricos em numéricos, necessário para o algoritmo CLAS. Todos são de problemas de classificação binária. Segue a descrição da seleção: *Statlog Australian Credit Approval* (australian); *Banknote Authentication* (banknote); *Breast Cancer Wisconsin* (breastcancer); *Breast Cancer Hess Probes* (breastHess); *Liver Disorders* (bupa); *Climate Model Simulation Crashes* (climate); *Pima Indian Diabetes* (diabetes); *Fertility* (fertility); *Statlog German Credit Data* (german); *Gene Expression* (golub); *Haberman's Survival* (haberman); *Statlog Heart Disease* (heart); *Indian Liver Patient* (ILPD); *Parkinsons* (parkinsons); *Connectionist Bench Sonar, Mines vs. Rocks* (sonar).

A Tabela I mostra as principais características dessa seleção, η representa a proporção entre as classes. Note a alta diversidade dos problemas para atestar a generalização do método.

B. Avaliação

Avaliou-se a métrica de qualidade Q média no espaço latente de uma rede neural em duas situações diferentes. Primeiro, variando a capacidade da rede, representada pelo número de neurônios na camada escondida, e, segundo, variando a regularização, a partir do parâmetro λ da norma L2 na função de custo com uma rede com número de neurônios alto, tentativa de provocar *overfitting*.

Tabela I
CARACTERÍSTICAS DAS BASES DE DADOS SELECIONADAS.

Dataset	Amostras	Atributos	η
australian	690	14	0.44
banknote	1372	4	0.44
breastcancer	683	6	0.65
breastHess	133	30	0.74
bupa	345	6	0.42
climate	540	18	0.91
diabetes	768	8	0.65
fertility	100	9	0.12
german	1000	24	0.70
golub	72	50	0.65
haberman	306	3	0.74
heart	270	13	0.56
ILPD	579	10	0.72
parkinsons	195	22	0.75
sonar	208	60	0.47

IV. RESULTADOS E DISCUSSÃO

A. Capacidade

O comportamento da métrica variando a capacidade do modelo foi muito parecido para todos os conjuntos de dados.

A Figura 1 mostra uma curva característica da avaliação no conjunto de treinamento.

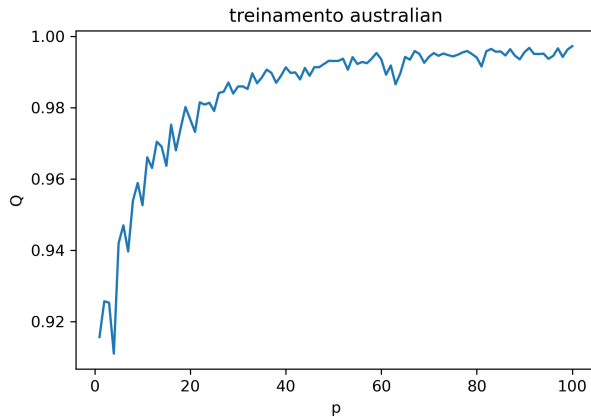


Figura 1. Evolução da métrica de qualidade média Q avaliando p no conjunto de treinamento; *dataset* australian.

Como previsto, quanto maior a capacidade do modelo maior também é a separabilidade dos dados o que é observado com o crescimento da métrica.

Contudo, um fenômeno parecido quebra um pouco a expectativa avaliando a métrica no conjunto de teste, como na Figura 2.

Uma esperança é que com o aumento da capacidade o modelo se especifique no conjunto de treinamento e a representação no conjunto de teste seja prejudicada. Contudo, ainda é observado o crescimento da métrica com o aumento da capacidade.

Uma explicação para esse fenômeno é um crescimento intrínseco da métrica para espaços com maior dimensão. O volume relativo de uma hipersfera, regra de formação do



Figura 2. Evolução da métrica de qualidade média Q avaliando p no conjunto de teste; *dataset* breastHess.

grafo, cai com o aumento da dimensão. Com mais espaço "vazio", a métrica tende a ser maior.

B. Regularização

Também, o comportamento da métrica variando agora a regularização do modelo foi bastante parecida para todos os conjuntos de dados.

O fenômeno também seguiu o que previa a expectativa, quanto maior a influência da regularização menor é a separabilidade dos dados representado pela diminuição da métrica.

A mesma curva característica foi observada no conjunto de treinamento e teste, como nas Figuras 3 e 4.

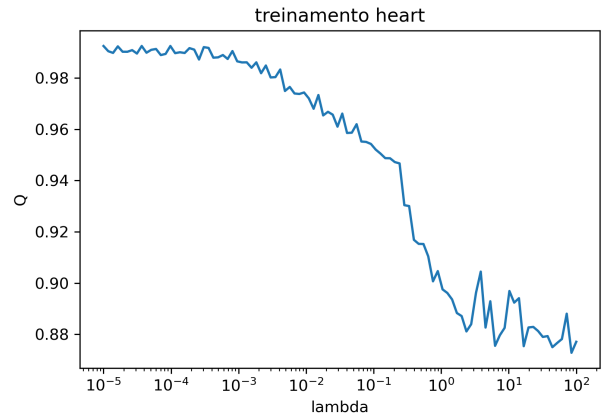


Figura 3. Evolução da métrica de qualidade média Q avaliando λ no conjunto de treinamento; *dataset* australian.

V. CONCLUSÕES

Este trabalho apresentou um estudo da representação latente aprendida de uma rede neural nas situações práticas de treinamento, capacidade e regularização, a partir de uma métrica de sobreposição baseada no Grafo de Gabriel.

Como prevê a teoria, a separabilidade, ou a sobreposição, do conjunto de dados aumentou com o aumento da capacidade e

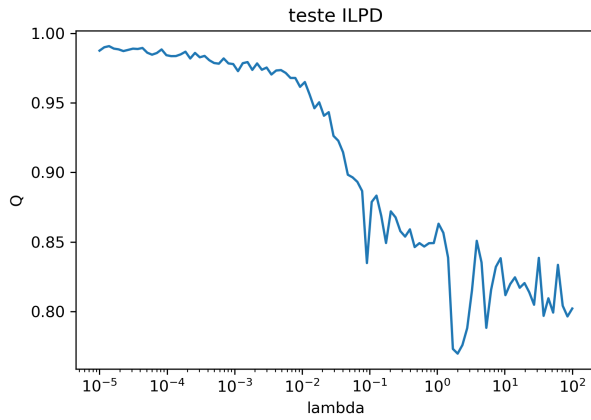


Figura 4. Evolução da métrica de qualidade média Q avaliando λ no conjunto de teste; *dataset* *australian*.

diminuiu com o peso da regularização. Um aspecto inesperado o comportamento da avaliação no conjunto de teste variando a capacidade, mas existe uma relação intrínseca entre a avaliação por distâncias da hipersfera e a dimensão do espaço.

AGRADECIMENTO

Este trabalho foi possível pela disponibilização das bases de dados pelo repositório *UCI Machine Learning* [4].

REFERÊNCIAS

- [1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] R. Salakhutdinov and G. Hinton, “Learning a nonlinear embedding by preserving class neighbourhood structure,” in *Artificial Intelligence and Statistics*, pp. 412–419, PMLR, 2007.
- [3] K. R. Gabriel and R. R. Sokal, “A new statistical approach to geographic variation analysis,” *Systematic zoology*, vol. 18, no. 3, pp. 259–278, 1969.
- [4] D. Dua and C. Graff, “UCI machine learning repository,” 2017.