

Lista 3 - MAC 0460
Bruno Mazetti Saito - 11221838

1 - Um problema de regressão logística seria resolvido por meio de uma equação com a forma $y_i = c_0 + c_1x_i + c_2 + x_i^2 \dots + c_nx_i^n (i = 1, 2, \dots, n)$, assim, a partir do conjunto de dados utilizados, podemos encontrar os coeficientes resolvendo um sistema de equações ou usando álgebra linear, por exemplo, e assim obter a equação correspondente ao ajuste a ser feito.

2 - Uma diferença notável é que o SVM busca encontrar a melhor superfície de decisão evitando, de certa forma, um possível overfitting, enquanto a regressão logística pode encontrar diferentes hiperplanos que podem ou não ser a solução ótima, além disso, a classificação do SVM é dada pelos próprios rótulos das classes, enquanto a regressão linear providencia as probabilidades referentes à cada classe. Em relação às semelhanças, podemos citar o fato de que ambos os algoritmos utilizam um método de erros de mínimos quadrados, e também o fato de que ambos apresentam um resultado semelhante, quando utilizado um *kernel* linear no SVM.

3 - Um ponto interessante a destacar sobre as redes neurais, é a sua capacidade de representar qualquer função contínua utilizando apenas uma camada oculta, outra característica a ser destacada é a utilização de várias camadas como perceptrons para poder fazer diferentes separações para cada uma delas, possibilitando realizar uma classificação mais exata. Uma qualidade interessante do SVM é a flexibilidade em relação aos possíveis dados ruidosos presentes no conjunto de treinamento, além disso, também podemos considerar a utilização de *kernel tricks* que são capazes de realizar uma transformação nos dados para maiores dimensões de forma mais eficiente e menos custosa.

4 - A validação serve basicamente como uma ajuda para o conjunto de treinamento, já que por meio dela podemos escolher os melhores hiperparâmetros para o modelo em questão afim de obter um bom resultado além de nos dar uma indicativa de *overfitting*. O teste, diferente da validação, não interfere no processo de treinamento, já que serve para se poder ter uma estimativa de como o modelo ajustado irá performar em um outro conjunto de dados.

5 - Ao meu entendimento, *overfitting* se trata de um problema em que o algoritmo "aprende de mais", ou seja, se um conjunto de treinamento possuir muitos dados ruidosos, o processo de aprendizagem pode ser comprometido ao tentar ajustar o modelo de forma que estes ruídos sejam ajustados corretamente, resultando em um ajuste não muito bom para outros conjuntos de dados não usados no treinamento. Desse modo, podemos detectá-lo por meio do conjunto de teste, já que se o modelo apresentar uma precisão muito alta para o conjunto de treinamento mas, em compensação, uma não tão alta para o conjunto de teste, isso nos dá uma clara evidência de *overfitting*. Para combatê-lo podemos utilizar a técnica de validação, utilizando, por exemplo, diversos subconjuntos

do conjunto de treinamento para que possamos ter um controle maior sobre os possíveis sobreajustes.

6 - • Acredito que dos tópicos destacados consegui ter um entendimento razoável em cerca de 75%;

- Um assunto que tive um entendimento considerável foi o de overfitting. O processo de overfitting é, basicamente, um problema encontrado nos algoritmos de treinamento que acabam minimizando o erro dentro da amostra mais do que deveriam, o que acaba resultando em um ajuste não muito bom para os dados fora do conjunto de treinamento;

- Sobre os assuntos de *clustering* e redes neurais convolucionadas não tive um avanço no conhecimento sobre eles principalmente pelo acúmulo de atividades do final do semestre, então acabei não vendo muito sobre estes assuntos;

- Na minha opinião, considero que tive um aproveitamento da disciplina de 7.5, apesar de não ter acompanhado alguns assuntos com a intenção que pretendia, aumentei o interesse pela área e gostei de implementar os algoritmos nos EP's.

7 - • Acabei estudando e procurando a respeito cerca de 80% dos conteúdos cobertos na disciplina;

- Realizei todas as atividades propostas;

- Assisti à grande maioria das aulas ao vivo, e quando não pude, procurei assistir às gravações depois. Consideraria que tive uma presença de 85%.

8 - Como dito no itens anteriores, tive um entendimento da maior parte dos conteúdos cobertos e assisti à grande parte das aulas também, embora não acabe participando muito com perguntas por preferência, além disso, acho que tive um bom desempenho nas tarefas propostas. No início do semestre esta era a disciplina que estava mais empolgado em cursar, e a implementação dos EP's aumentou mais o meu interesse pela área, e inclusive, me fez considerar a realização do TCC na área.