Anotações do Livro Elements of Statistical Learning - Cap 02: Overview of Supervised Learning

Rafael Barbosa da Silva

30 de maio de 2020

1 Introdução

- Nos exemplos anteriores do livro, tem-se variáveis chamadas de *inputs*, isto é, variáveis de entrada (explicativas, independentes, features, etc) e alguma(s) variáveis chamadas de *output*, chamada de target, dependente, etc.

2 Tipos de variáveis e terminologia

- Dependendo da natureza do problema, podemos ter um tipo de variável para a target;
- Na predição de glicose, esta variável Y foi quantitativa;
- No dataset Iris, a variável Y é qualitativa com 3 categorias (Virginica, Setosa e Versicolor);
- Podemos prever estas variáveis Y a partir de características (variáveis X) do fenômeno em que elas estão encaixadas;
- Exemplo: Dado medidas atmosféricas de ontem e hoje, queremos prever o nível de Ozônio amanhã;
- Nas tarefas em que a variável que queremos prever é quantitativa, chamamos o fenômeno de regressão;
- Nas tarefas em que a variável que queremos prever é qualitativa, chamamos o fenômeno de classificação;
- Ambas podem ser vistas como uma tarefa de aproximação;
- As variáveis X também podem ter natureza quantitativa ou qualitativa e cada modelo/método de previsão pode ter sua preferência de variável;
- Existe também a variável ordinal, que possuem categorias em uma ordem. Exemplo, escolaridade: ensino fundamental, médio e superior;
- Variáveis qualitativas podem ser codificadas para quantitativas, elas melhoram o desempenho dos modelos computacionalmente. Exemplo 0: Não sobreviveu, 1: Sobreviveu;

- Se possível, leia sobre variáveis Dummies.

3 Duas abordagens de predição: Mínimos quadrados e Vizinhos mais Próximos

- São duas abordagens robustas de previsão;
- A partir desse momento entenda que as variáveis X podem ser chamadas de **explicativa/features/dependentes** e a Y pode ser chamada de **resposta/ target/independente**;

3.1 Modelos lineares e Mínimos Quadrados

- Dado que temos um vetor de features $X^T = (X_1, X_2, \ldots, X_p)$, podemos prever a variável target Y pelo modelo:

$$\hat{Y} = \hat{\beta}_0 + \sum_{j=1}^{p} X_j \hat{\beta}_j$$

- O termo $\hat{\beta}_0$ é chamado de intercepto, conhecido no ML como *bias*;
- Podemos escrever essa equação como um produto de matrizes:

$$\hat{Y} = X^T \hat{\beta}$$

- Queremos escolher os melhores Betas que minimizam a soma dos quadrados dos resíduos:

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - x_i^T \beta)^2$$

- Resumindo a demonstração, chegamos na fórmula abaixo para estimar os Betas:

$$\hat{Y} = (\boldsymbol{X^T X})^{-1} \boldsymbol{X^T y}$$

- **Exemplo:** Na Figura 2.1, ele mostra um scatterplot e nos diz sobre um modelo linear de classificação. A variável resposta Y, nesse caso G, possui duas categorias Blue e Orange;
- Em cada uma das categorias temos 100 observações;
- Uma de regressão linear foi modelada para esses dados, com a variável resposta Y recodificada como 0: Blue e 1: Orange;
- E segundo a regra abaixo:
- Se o valor estimado $(\hat{Y}) > 0.5$, então a categoria é 1 (Orange).
- Se o valor estimado $(\hat{Y}) \leq 0.5$, então a categoria é 0 (Blue);