SME0305 - 2014

Gustavo Carlos Buscaglia

ICMC - Ramal 738176
gustavo.buscaglia@gmail.com

Lista 8 (28 de abril de 2014)

---

1. Primeiramente, assistir os vídeos atentamente e responder perguntas tais como
   (a) O que é um método iterativo?
   (b) Quando se diz que um método iterativo "chegou" à solução?
   (c) Quando chega a uma solução, podem existir outras?
   (d) Como se sabe se um método de minimização chegou a um mínimo da função? Conceito de "tolerância".
   (e) Tipo de problemas de classificação e de regressão.
   (f) Concretamente, o que é "treinar" um método de regressão linear?
   (g) Como é a função hipótese quando se tem muitas características (variáveis)? Quantos coeficientes indeterminados tem?
   (h) Como são as linhas de custo constante no plano $\theta_0 - \theta_1$?
   (i) Qual o efeito sobre essas linhas de fazer "scaling" das variáveis?

2. No caso de escolher a função custo como
$$\widetilde{J}(\theta_0, \theta_1) = \frac{1}{6\,m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^6 \,,$$
o par de valores de (θ_0, θ_1) que minimiza $\widetilde{J}(\theta_0, \theta_1)$ é idêntico àquele que minimiza
$$J(\theta_0, \theta_1) = \frac{1}{2\,m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2 \ \ ?$$

3. Refazer os cálculos contidos nos slides de Andrew Ng (comentados em sala), mas considerando a função custo $\widetilde{J}(\theta_0, \theta_1)$ (com conjunto de treino $\{(1, 1), (2, 2), (3, 3)\}$ e h_θ como uma função linear: $h_\theta(x) = \theta_1 x$) – ver vídeo `0203costfunctioni.mp4` ao tempo `3min00s`.

   Dica: pensar na condição necessária para que uma função de duas variáveis admita um mínimo local.

4. Provar que as derivadas da função
$$J(\theta_0, \theta_1) = \frac{1}{2\,m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)^2 \,,$$
são
$$\frac{\partial J}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)$$
e
$$\frac{\partial J}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta(x_i) - y_i\right)\, x_i$$
Lembre que $h_\theta(x) = \theta_0 + \theta_1\, x$.

5. Provar que, definindo o vetor $\underline{X}$ como o vetor de componentes x_1, x_2, ..., x_m, definindo o vetor $\underline{Y}$ analogamente, e o vetor $\underline{\mathbf{1}}$ com todas as componentes iguais a 1, se cumpre que
$$J(\theta_0, \theta_1) = \frac{1}{2m}\, \|\theta_0\, \underline{\mathbf{1}} + \theta_1\, \underline{X} - \underline{Y}\|_2^2$$
Obs: Analize as operações algébricas que estão simbolizadas pelas operações vetoriais e pela sumatória, e ver que não há nenhuma diferença.

6. Provar ainda que, usando a notação anterior,
$$\frac{\partial J}{\partial \theta_0} = \frac{1}{m}\, (\theta_0\, \underline{\mathbf{1}} + \theta_1\, \underline{X} - \underline{Y})^T\ \underline{\mathbf{1}}$$
e que
$$\frac{\partial J}{\partial \theta_1} = \frac{1}{m}\, (\theta_0\, \underline{\mathbf{1}} + \theta_1\, \underline{X} - \underline{Y})^T\ \underline{X}$$

7. Comparar a função J com a função minimizada quando se resolve o sistema incompatível (ver livro)
$$\underline{\underline{A}}\, \underline{\theta} = \underline{Y} \qquad\qquad (*)$$
sendo $\underline{\underline{A}}$ a matriz com (duas) colunas, $\underline{\mathbf{1}}$ e $\underline{X}$, e $\underline{\theta} = (\theta_0, \theta_1)^T$.

   Concluir que a regressão linear é equivalente à resolução por mínimos quadrados do sistema linear $(*)$.

8. Provar que
$$J(\theta_0, \theta_1) = \frac{1}{2m} \left(\underline{\underline{A}}\, \underline{\theta} - \underline{Y}\right)^T \left(\underline{\underline{A}}\, \underline{\theta} - \underline{Y}\right)$$
cujo gradiente no plano θ_0-θ_1 vale
$$\nabla J\,(\underline{\theta}) \ = \ \begin{pmatrix} \frac{\partial J}{\partial \theta_0}(\underline{\theta}) \\ \\ \frac{\partial J}{\partial \theta_1}(\underline{\theta}) \end{pmatrix} \ = \ \frac{1}{m} \left(\underline{\underline{A}}^T\, \underline{\underline{A}}\, \underline{\theta} - \underline{\underline{A}}^T\, \underline{Y}\right)$$
Notar que o fator $1/m$ não muda o valor de $\underline{\theta}$ que faz zero o gradiente, isto é, o mínimo.

9. Sabendo que
$$J(\underline{\theta}) = \frac{1}{2m} \left(\underline{\underline{A}}\, \underline{\theta} - \underline{Y}\right)^T \left(\underline{\underline{A}}\, \underline{\theta} - \underline{Y}\right)$$
e sendo $\underline{d} = (d_0, d_1)^T$ a direção na qual se quer atualizar $\underline{\theta}$, isto é
$$\underline{\theta}^{new} = \underline{\theta}^{old} + \alpha\, \underline{d}$$
calcular os coeficientes do polinômio
$$p(\alpha) = a + b\, \alpha + c\, \alpha^2$$
tais que
$$p(\alpha) = J\left(\underline{\theta}^{old} + \alpha\, \underline{d}\right)$$
Notar que esses coeficientes são função de $\underline{\theta}^{old}$, de $\underline{d}$, de $\underline{\underline{A}}$, de $\underline{Y}$, e de m, mas não de α. Notar por exemplo que
$$a = \frac{1}{2m}\, (\underline{\underline{A}}\, \underline{\theta}^{old} - \underline{Y})^T\, (\underline{\underline{A}}\, \underline{\theta}^{old} - \underline{Y})$$

10. Pensar se a seguinte afirmação é ou não verdadeira: "Considerando-se a minimização da diferença quadrática entre o valor da variável alvo e seu respectivo valor predito pela regressão, se forem invertidas as variáveis *input* e *output* ('alvo') de um certo problema de interesse, a minimização da função custo resultará na mesma função afim daquela que é obtida sem se fazer a inversão". Por exemplo, em relação a "estimar o preço de venda de um imóvel através de sua área construída" tal inversão resultaria em "estimar a área construída de um imóvel que foi vendido por um dado preço conhecido".

11. No caso da afirmação do exercício anterior ser verdadeira (respectivamente, falsa), em qual ou quais casos ou situações ela torna-se falsa (respectivamente, verdadeira)?

    Dica: pense em conjuntos treino específicos e em outras noções de distância para a função custo (i.e., diferentes daquela contida no exercício anterior).

12. (Opcional) Tendo resolvido o exercício 1, utilize o Octave ou o MATLAB para gerar um conjunto de dados $\{(x_i, y_i)\}$ de tamanho 20 ($i = 1, 2, \ldots, 20$), em que y_i deve ser construído como uma perturbação da função afim $f(x) = -5\,x + 3$ somando-se a esta um número aleatório proveniente da distribuição normal padrão (média zero e desvio padrão 0.1). Salve tal conjunto de dados em um arquivo e depois construa um código que através da leitura de tal arquivo calcule o valor do par (θ_0, θ_1) que minimiza as funções J e $\widetilde{J}$ definidas no exercício 1. Para comparação, calcule também o ajuste linear utilizando diretamente a função `polyfit`.

Boa prática!!