Processamento Estatístico da Linguagem Natural

Aula 15

Professora Bianca
(Sala 302 – Bloco E)
bianca@ic.uff.br
http://www.ic.uff.br/~bianca/peln/

Aula de Hoje

- Cap. 6 – Jurafsky & Martin – Hidden Markov and Maximum Entropy Models
  - Seção 6.6

Modelos de Entropia Máxima

- A segunda abordagem para classificação sequencial (primeira = HMM) que veremos é o "**Modelo de Markov de Entropia Máxima**" (MEMM).
- Antes temos que aprender sobre o modelo de **classificação não-sequencial** no qual ela é baseada.
  - Entropia Máxima = MaxEnt = Regressão Logística Multinomial.

Classificação (não-sequencial)

- Na tarefa de classificação:
  - Recebemos como entrada um único objeto.
  - Extraímos características (ou atributos) importantes desse objeto.
  - Classificamos o objeto de acordo com as suas características em um conjunto discreto de classes.
    - Em muitos casos obtem-se uma distribuição de probabilidade sobre as classes.
- Exemplos:
  - Classificação de spam vs. não-spam.

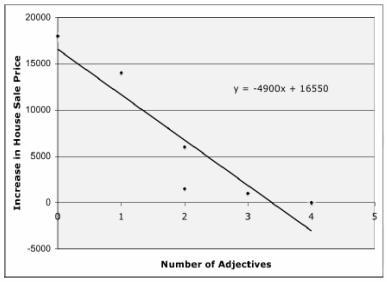
Classificador MaxEnt

- Combina os valores dos atributos linearmente e usa uma exponencial para determinar a classe.
  - O aprendizado do classificador consiste em encontrar os pesos da combinação linear.
- Para entender esse classificador precisamos primeiro ver dois métodos clássicos:
  - Regressão linear
  - Regressão logística

Regressão Linear

- Regressão: previsão de um valor real e não de uma classe.
- Dados de treinamento: conjunto de observações (x,y) onde x é um atributo do objeto e y é o valor a ser previsto.
- O modelo é a reta y = ax + b que melhor se encaixa nas observações.
  - Este modelo tem dois parâmetros:
    - a é a inclinação da reta
    - b é a interseção da reta com o eixo y.

Exemplo: Regressão Linear



De forma mais geral

- Podemos usar mais de um atributo na regressão linear.
  - Regressão linear múltipla.
    - $f_1..f_N$ = vetor de atributos
    - $w_1..w_N$ = parâmetros = vetor de pesos
    - w_0 = interseção do hiperplano com o eixo y

$$y = w_0 + \sum_{i=1}^{N} w_i f_i$$

  - Para simplificar podemos usar um atributo f_0 com valor fixo em 1 e escrever a equação como:

$$y = \mathbf{w} \cdot \mathbf{f}$$

Aprendizado em Regressão Linear

- Minimizar a soma dos erros quadráticos para todos os exemplos de treinamento:

$$cost(W) = \sum_{j=0}^{M} (y_{pred}^{(j)} - y_{obs}^{(j)})^2$$

onde

$$y_{pred}^{j} = \sum_{i=0}^{N} w_i f_i^{(j)}$$

- Fórmula fechada para escolher o conjunto de pesos W que minimiza o erro é:

$$W = (X^T X)^{-1} X^T \overrightarrow{y}$$

onde a matrix X contém os atributos de todos os exemplos de treinamento e o vetor y contém os valores de y_{obs}.

Regressão Logística

- Na regressão logística, usamos o modelo linear para fazer classificação.
  - Modelo linear que atribui probabilidades às classes.
  - Para classificação binária, queremos prever P(y=true|x).
- Na regressão linear o valor previsto é um valor do domínio dos números reais, aqui queremos prever um valor entre 0 e 1.

Regressão Logística

- Para transformar um valor entre 0 e 1 em um valor entre -∞ e +∞ usamos a transformação logística:

$$ln\left(\frac{p(y = true|x)}{1 - p(y = true|x)}\right) = \mathbf{w} \cdot \mathbf{f}$$

$$\frac{p(y = true|x)}{1 - p(y = true|x)} = e^{\mathbf{w} \cdot \mathbf{f}}$$

- Resolvendo para $p(y = true|x)$, obtemos:

$$p(y = true|x) = \frac{e^{\mathbf{w} \cdot \mathbf{f}}}{1 + e^{\mathbf{w} \cdot \mathbf{f}}}$$

$$= \frac{exp\left(\sum_{i=0}^{N} w_i f_i\right)}{1 + exp\left(\sum_{i=0}^{N} w_i f_i\right)}$$

Classificação com Regressão Logística

- Exemplo *x* pertence à classe *true* se:

$$\frac{p(y = true|x)}{1 - p(y = true|x)} > 1$$

$$e^{\mathbf{w} \cdot \mathbf{f}} > 1$$

$$\mathbf{w} \cdot \mathbf{f} > 0$$

$$\sum_{i=0}^{N} w_i f_i > 0$$

- A equação $\sum_{i=0}^{N} w_i f_i = 0$ define um **hiperplano** de separação entre as classes.

Aprendizado em Regressão Logística

- Estimativa de probabilidade condicional.
  - Escolher os pesos que fazem a probabilidade dos valores observados y serem as maiores, dadas as observações x.
  - Para um conjunto de treinamento com m exemplos:

$$\hat{\mathbf{w}} = \underset{w}{\operatorname{argmax}} \prod_{i=0}^{M} P(y^{(i)}|x^{(i)})$$

  - Resolver essa equação com técnicas de otimização convexa (L-BFGS, subida de gradiente, etc.)

Modelo de Entropia Máxima

- Regressão logística com mais de 2 classes = regressão logística multinomial.
  - Também conhecida como Modelo de Entropia Máxima (MaxEnt).
  - A equação do MaxEnt é:

$$p(c|x) = \frac{exp\left(\sum_{i=0}^{N} w_{ci} f_i\right)}{\sum_{c' \in C} exp\left(\sum_{i=0}^{N} w_{c'i} f_i\right)}$$

  - O denominador é um fator de normalização comumente chamado de Z:

$$p(c|x) = \frac{1}{Z} exp \sum_{i=0}^{N} w_{ci} f_i$$