

Projeto de Aprendizagem de Máquina

Dayvid Victor Rodrigues de Oliveira e Denys Lins Farias

12 de maio de 2012

Resumo

ABSTRACT

Sumário

1	Introdução	1
1.1	Conceito	1
1.2	Objetivo	1
1.3	Estrutura do Trabalho	1
2	Técnicas Avaliadas	2
2.1	Contexto e Histórico	2
2.1.1	Sistemas de Reconhecimento de Padrões	2
2.2	Classificadores	2
2.2.1	K-Nearest Neighbor	3
2.2.2	Janela de Parzen	3
2.2.3	Estimação da Máxima Verossimilhança	4
2.2.4	Algoritmo Expectation-Maximization	4
2.3	Combinação de Classificadores	4
2.3.1	Regra do Produto	4
2.3.2	Regra da Soma	5
2.3.3	Regra do Máximo	5
2.4	Regra do Mínimo	5
2.4.1	Regra da Mediana	5
2.5	Agrupamento	5
2.5.1	K-Means	6
3	Experimentos	7
3.1	Metodologia	7
3.2	Geração de Dados	7
3.3	Avaliação	7
3.4	Comparação	7
4	Conclusão	8
	Referências Bibliográficas	9

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

1.1 Conceito

1.2 Objetivo

1.3 Estrutura do Trabalho

Capítulo 2

Técnicas Avaliadas

2.1 Contexto e Histórico

Esta sessão abordará sistemas de reconhecimento de padrões e classificadores com aprendizagem baseadas em instâncias.

2.1.1 Sistemas de Reconhecimento de Padrões

No final dos anos 50, surgiram os primeiros trabalhos de aprendizagem de máquina. De uma forma geral, elas consistiam em dar ao computador a habilidade de reconhecer formas. A partir daí, surgiram diversos problemas onde a aprendizagem de máquina atuava.

Existem três problemas gerais que a aprendizagem de máquina tenta resolver. Um deles é o problema do agrupamento, que consiste em agrupar dados de acordo com suas características, de forma que seja possível extrair informação útil destes agrupamentos. Um outro problema é a discriminação, que basicamente é achar uma forma de reconhecer um conceito, dado um conjunto de conceitos exemplos. O terceiro e último problema é o da generalização, que é o problema de como reduzir uma regra de classificação, tornando-a mais abrangente e menos custosa.

Reconhecimento de padrões ataca principalmente o problema da discriminação, tendo por objetivo classificar padrões, discriminando-os entre duas ou mais classes. A classificação pode ser feita com padrões pertencentes a qualquer domínio, como reconhecimento de digitais, gestos, escrita, fala, entre outros.

2.2 Classificadores

Todo sistema de reconhecimento de padrões utiliza um classificador para discriminar os padrões de teste. A eficiência de um classificador é medida pela taxa de acerto média, pela variância, e pelo seu custo computacional. Um classificador de aprendizagem baseada em instâncias muito utilizado é o *K-Nearest Neighbor*, KNN [Patrick and II (1969)].

2.2.1 K-Nearest Neighbor

O KNN é muito usado por ser um método de aprendizagem supervisionado simples, e por possuir uma taxa de acerto relativamente alta. O conceito básico consiste em: Dado um padrão x a ser classificado e um conjunto de padrões conhecidos T , obter as classes dos K elementos de T mais próximos de x . A classe que obtiver maior ocorrência, ou peso, será a classe de x . Pode-se dizer que o KNN utiliza uma abordagem "*Dize-me com quem andas, e direi quem és.*". O algoritmo está descrito em Algorithm 1.

Algorithm 1 KNN

Require: K : um número

Require: T : conjunto de treinamento

Require: x : elemento para ser classificado

Require: L : uma lista

1: **for all** $t_i \in T$ **do**

2: $d_i = \text{distance}(t_i, x)$

3: adicione $(d_i, \text{Classe}(t_i))$ em L

4: **end for**

5: $\text{Ordene}(L)$ de acordo com as distâncias

6: obtenha os K primeiros elementos de L

7: **return** a classe de maior ocorrência, ou peso, entre os K

Pode-se dizer que o K – *NearestNeighbor* atua considerando a densidade das classes na região onde o padrão que se deseja classificar. A estimação de densidade é baseada na probabilidade a posteriori.

2.2.2 Janela de Parzen

Janela de Parzen é uma técnica de estimação de densidade que se baseia na interpolação de dados. Esta estimativa é feita assumindo-se uma função $K(x)$ que determina a janela centrada em x com largura h , esta função é chamada de função de kernel.

De maneira simples, para estimar a densidade fixa-se uma região R e um volume fixo V , e k correspondente é determinado a partir dos dados de aprendizagem. Assumindo que a região R é um hipercubo de d dimensões, o seu volume é dado por h^d .

$$p(x) = \frac{k}{V} \quad (2.1)$$

Para se obter a função de densidade de probabilidade de uma amostra x utilizando-se a Janela de Parzen, simplesmente centra-se a região R em x e conta-se o número de instâncias dentro desta região (valor de k) na equação 2.1.

Determinar o h ideal é uma tarefa minuciosa, pois, se o valor de h for muito pequeno, haverá muita especialização e estará mais sujeita a erros ocasionados por ruídos, já se h for muito grande, ocorrerá uma super-generalização. A escolha da função $K(x)$ apropriada também é um problema para esta técnica.

2.2.3 Estimação da Máxima Verossimilhança

Estimativa por máxima verossimilhança, *Maximum – Likelihood*, é um método para estimar os parâmetros de um modelo estatístico. A partir de um conjunto de dados e um dado modelo estatístico (i.e. Normal), o *MLE* estima valores adequados para os parâmetros do modelo.

De forma geral, dado um conjunto de instâncias de treinamento e um modelo estatístico, o *MLE* estima os valores dos parâmetros do modelo estatístico de forma que a probabilidade dos dados observados sejam maximizados.

2.2.4 Algoritmo Expectation-Maximization

Expectation-Maximization é um algoritmo iterativo para se encontrar a máxima verossimilhança, os parâmetros de um modelo, onde o modelo é desconhecido, e depende de variáveis não observadas.

A iteração do *EM* alterna entre realizar a Expectation *E* e a Maximization *M*. A *E* computa o valor esperado da log-verossimilhança avaliada, utilizando a estimativa atual para os parâmetros. Já a *M* computa os parâmetros, maximizando a esperada log-verossimilhança encontrada em *E*.

A diferença do *EM* para o *MLE* é que o *EM* é capaz de tratar distribuições multimodais, além de ser capaz de estimar os parâmetros mesmo com *missingvalues*.

2.3 Combinação de Classificadores

Combinar classificadores é a ação de utilizar dois classificadores para classificar um dado padrão. No caso, a ideia é obter o melhor de cada classificador. A combinação de classificadores deve ser feita utilizando classificadores que diferem, que classificam um mesmo padrão de forma divergente.

Por combinar "o melhor" dos classificadores, esta técnica é interessante para tratar de regiões de indecisão, driblando limitações dos classificadores utilizados, pois cada um fornece uma informação complementar na tarefa de classificar um padrão.

Existem diversas formas de combinar classificadores, algumas delas serão abordadas logo abaixo.

2.3.1 Regra do Produto

Uma forma de fazer a combinação de classificadores é utilizando a regra do produto. Esta regra, como o próprio nome sugere, nada mais é do que fazer a multi-

plicação das probabilidades *a posteriori* de cada classificador.

Dado um padrão x , ele será classificado como sendo da classe que obtiver maior valor da multiplicação entre as probabilidades *a posteriori* de um classificador.

Esta técnica é adequada quando as entradas dos classificadores são providas por sensores diferentes.

2.3.2 Regra da Soma

A regra da soma faz a combinação de classificadores somando as probabilidades *a posteriori* de cada classificador. Dado um padrão x , ele será classificado como sendo da classe que obtiver maior soma de probabilidade.

A regra da soma é adequada para classificadores que possuem o mesmo sensor, pois o mesmo não é muito sensível a erros na estimação de densidade. Para se obter melhores resultados, pode-se atribuir diferentes pesos para as probabilidades *a posteriori* de cada classificador.

2.3.3 Regra do Máximo

A regra do máximo é uma aproximação da regra da soma. Nesta aproximação, a probabilidade de uma determinada classe é a maior probabilidade *a posteriori* dentre as saídas dos classificadores para aquela classe.

2.4 Regra do Mínimo

A regra do mínimo é uma aproximação da regra do produto. Nesta aproximação, a probabilidade de uma determinada classe é a menor das probabilidades *a posteriori* dentre as saídas dos classificadores para aquela classe.

2.4.1 Regra da Mediana

Esta regra é um caso particular onde é obtida as medianas das probabilidades *a posteriori* de cada classe. A classe que obtiver maior mediana é utilizada para classificar um padrão x de entrada.

2.5 Agrupamento

Agrupamento, também conhecido como Clusterização, é a classificação não-supervisionada de dados que forma grupos, ou clusters. Assim, a análise de clusters envolve a organização de um conjunto de padrões em agrupamentos, de acordo com algumas regras de similaridade.

De forma simples, pode-se dizer que a clusterização é o agrupamento de instâncias baseadas nas suas características, ou seja, instâncias similares pertencem ao mesmo grupo. Estes agrupamentos são formados por padrões não rotulados, ou

seja, não se sabe a classe de cada padrão, assim, rótulos podem ser definidos após a realização do agrupamento.

Uma vantagem da clusterização é que atualmente, existem muitos dados, mas poucos destes dados são classificados. Algoritmos de clusterização possibilitam a rotulação de forma mais rápida e factível.

2.5.1 K-Means

Um dos algoritmos de clusterização mais famosos é o K-means. Este método divide T padrões em K clusters onde cada padrão x pertence ao cluster de média mais próxima de x . O algoritmo é repetido até que poucas alterações sejam feitas em cada iteração. O resultado deste algoritmo é uma partição do espaço em Células de Voronoi.

Algorithm 2 K-Means

Require: K : um número de clusters desejados

Require: T : um conjunto de padrões

```
1:  $M$  = Obtenha  $K$  padrões aleatoriamente
2:  $G$  = Grupo de instâncias de cada padrão em  $M$ 
3: while  $M$  não convergiu do
4:   for all  $t_i \in T$  do
5:      $j$  = index de  $KNN(t_i, M)$ 
6:     adicione  $t_i$  ao grupo  $G_j$ 
7:   end for
8:   atualize  $M$  com as médias de  $G$ 
9: end while
10: return os clusters e as medias  $G, M$ 
```

Capítulo 3

Experimentos

3.1 Metodologia

3.2 Geração de Dados

3.3 Avaliação

3.4 Comparação

Capítulo 4

Conclusão

ADICIONAR CONCLUSÃO.

Referências Bibliográficas

Patrick, E. A. and II, F. P. F. (1969) A generalization of the k-nearest neighbor rule.
In *IJCAI*, pp. 63–64.