

# Projeto de Aprendizagem de Máquina

Dayvid Victor Rodrigues de Oliveira e Denys Lins Farias

9 de maio de 2012

## **Resumo**

## **ABSTRACT**

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Conceito . . . . .	1
1.2	Objetivo . . . . .	1
1.3	Estrutura do Trabalho . . . . .	1
<b>2</b>	<b>Técnicas Avaliadas</b>	<b>2</b>
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	Expectation-Maximization . . . . .	4
2.2.5	Combinação de Classificadores . . . . .	4
2.3	Agrupamento . . . . .	4
2.3.1	K-Means . . . . .	4
<b>3</b>	<b>Experimentos</b>	<b>5</b>
3.1	Metodologia . . . . .	5
3.2	Geração de Dados . . . . .	5
3.3	Avaliação . . . . .	5
3.4	Comparação . . . . .	5
<b>4</b>	<b>Conclusão</b>	<b>6</b>
	<b>Referências Bibliográficas</b>	<b>7</b>

# **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:    $d_i = distance(t_i, x)$ 
3:   adicione  $(d_i, Classe(t_i))$  em  $L$ 
4: end for
5:  $Ordene(L)$  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 Expectation-Maximization**

### **2.2.5 Combinação de Classificadores**

## **2.3 Agrupamento**

### **2.3.1 K-Means**

## **Capítulo 3**

# **Experimentos**

### **3.1 Metodologia**

### **3.2 Geração de Dados**

### **3.3 Avaliação**

### **3.4 Comparação**

## Capítulo 4

# Conclusão

Com os resultados observados, pode-se concluir que o ASGP e o ASGPM possuem uma redução de instâncias intermediária, entre o SGP1 e o SGP2.

Quanto a taxa de acerto, percebe-se que as adaptações propostas ocasionam uma pequena queda na taxa de acerto da classe majoritária, ocasionando uma queda também na taxa de acerto geral. Porém, observa-se que, na grande maioria dos casos, a percentagem de acerto ganha na classe minoritária é muito superior a percentagem perdida na classe majoritária.

Assim, conclui-se que o ASGP e o ASGPM são técnicas eficientes quando se deseja priorizar a classe minoritária, e é necessário reduzir drasticamente o conjunto de instâncias.

Para trabalhos futuros, propõe-se que sejam avaliadas novas formas de utilizar o *Merge* e o *Pruning*, para encontrar um equilíbrio entre a remoção de instâncias da classe majoritária e a representação da classe minoritária.

# Referências Bibliográficas

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