



浙江大学

ZheJiang University

支持向量机的应用问题

浙江大学信电学院 胡浩基

Haoji_hu@zju.edu.cn

13819176516



浙江大学

ZheJiang University

SVM内核总结

- **Linear**（线性内核）： $K(x, y) = x^T y$
- **Ploy**（多项式核）： $K(x, y) = (x^T y + 1)^d$
- **Rbf**（高斯径向基函数核）： $K(x, y) = e^{-\frac{\|x-y\|^2}{\sigma^2}}$
- **Tanh**（**Tanh** 核）：

$$K(x, y) = \tanh(\beta x^T y + b)$$

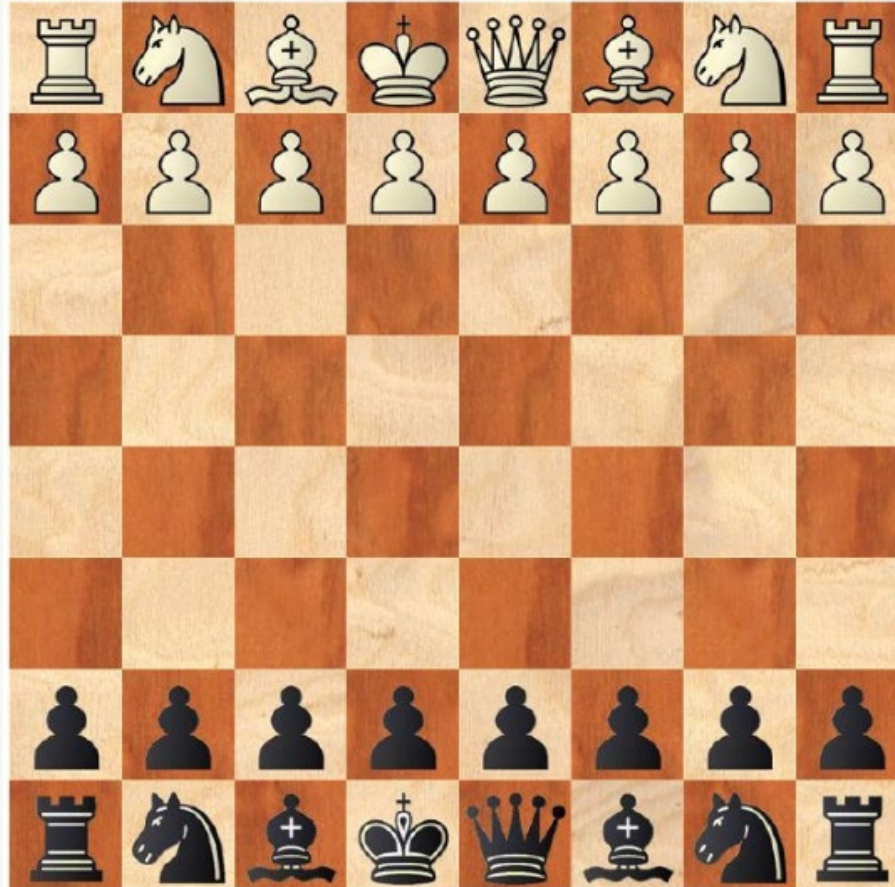
$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



浙江大学

ZheJiang University

国际象棋规则介绍





浙江大学

ZheJiang University

国际象棋规则介绍

- 兵（黑白各**8**个）：第一步向前可走一格或两格，以后每次只能向前走一格，不能后退。但在吃对方子时，则是向位于斜前方的那格去吃，并落在那个格。





浙江大学

ZheJiang University

国际象棋规则介绍

- 王（黑白各**1**个）：是国际象棋中最为重要的棋子，王被将死即告负。走法是每次横直斜走均可，但每次只能走一格。吃子与走法相同。





浙江大学
ZheJiang University

国际象棋规则介绍

- 兵王问题：黑方只剩一个王，白方剩一个兵一个王。
 - 两种可能
 - (1) 白方将死黑方，获胜。
 - (2) 和棋。
- 这两种可能视三个棋子在棋盘的位置而确定。

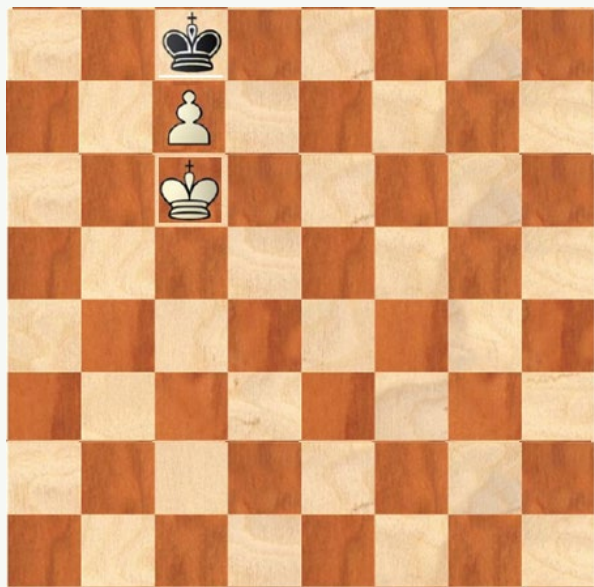


浙江大学

ZheJiang University

国际象棋规则介绍

- 兵的升变：兵走至对方底线，可以升变为除王以外的任意一子。
- 逼和：一方的王未被将军，但移动到任意地方都会被对方将死，则此时是和棋。





用SVM解兵王问题

■ UCI Machine Learning Repository

<http://archive.ics.uci.edu/ml/datasets.html>



Machine Learning Repository
Center for Machine Learning and Intelligent Systems






[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

☒ Repository ☐ Web

[View ALL Data Sets](#)

Browse Through: 394 Data Sets

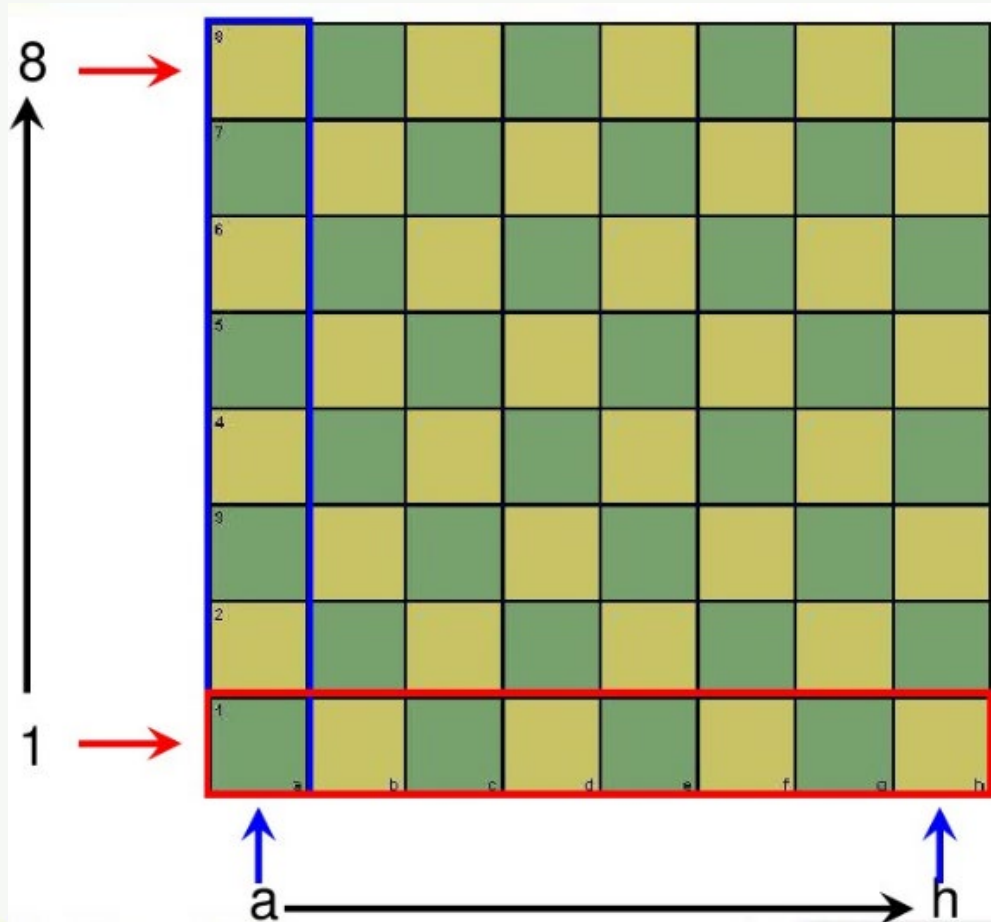
Table View [List View](#)

Default Task	Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
Classification (289) Regression (74) Clustering (67) Other (54)	 Abalone	Multivariate	Classification	Categorical, Integer, Real	4177	8	1995
Attribute Type Categorical (37) Numerical (244) Mixed (55)	 Adult	Multivariate	Classification	Categorical, Integer	48842	14	1996
Data Type Multivariate (306) Univariate (16) Sequential (40) Time-Series (75) Text (37) Domain-Theory (22) Other (21)	 Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Area Life Sciences (89)	 Anonymous Microsoft Web Data		Recommender-Systems	Categorical	37711	294	1998
	 Arrhythmia	Multivariate	Classification	Categorical, Integer, Real	452	279	1998



用SVM解兵王问题

■ krkopt.data文件



```
a,1,b,3,c,2,draw
a,1,c,1,c,2,draw
a,1,c,1,d,1,draw
a,1,c,1,d,2,draw
a,1,c,2,c,1,draw
a,1,c,2,c,3,draw
a,1,c,2,d,1,draw
a,1,c,2,d,2,draw
a,1,c,2,d,3,draw
a,1,c,3,c,2,draw
a,1,c,3,d,2,draw
a,1,c,3,d,3,draw
a,1,c,3,d,4,draw
. . .
d,1,e,3,f,1,six
d,1,e,3,g,1,six
d,1,e,3,g,2,six
d,1,e,4,h,1,six
d,1,e,4,h,2,six
d,1,e,4,h,3,six
d,1,e,5,h,1,six
d,1,e,6,h,1,six
. . .
c,1,e,7,c,5,fifteen
c,1,e,7,c,6,fifteen
c,1,e,7,c,7,fifteen
c,1,e,7,d,5,fifteen
c,1,e,7,e,5,fifteen
c,1,f,1,c,3,fifteen
```



浙江大学
ZheJiang University

用SVM解兵王问题

- **LIBSVM -- A Library for Support Vector Machines**
<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

LIBSVM -- A Library for Support Vector Machines

Chih-Chung Chang and [Chih-Jen Lin](#)

NEW Version 3.22 released on December 22, 2016. It conducts some minor fixes.

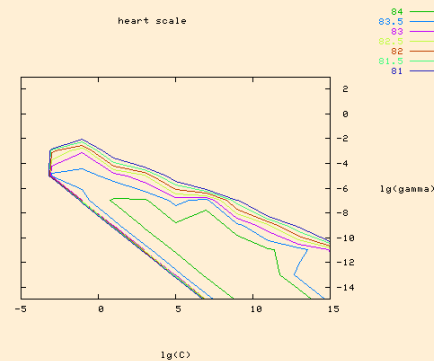
NEW [LIBSVM tools](#) provides **many extensions** of LIBSVM. Please check it if you need some functions not supported in LIBSVM.

NEW We now have a nice page [LIBSVM data sets](#) providing problems in LIBSVM format.

NEW [A practical guide to SVM classification](#) is available now! (mainly written for beginners)

We now have an easy script (easy.py) for users who know NOTHING about SVM. It makes everything automatic--from data scaling to parameter selection.

The parameter selection tool grid.py generates the following contour of cross-validation accuracy. To use this tool, you also need to install [python](#) and [gnuplot](#).



To see the importance of parameter selection, please see our [guide](#) for beginners.

NEW Using libsvm, our group is the winner of IJCNN 2001 Challenge (two of the three competitions), EUNITE world wide competition on electricity load prediction, [NIPS 2003 feature selection challenge](#) (third place), [WCCI 2008 Causation and Prediction challenge](#) (one of the two winners), and [Active Learning Challenge](#) 2010 (2nd place).



用SVM解兵王问题

- 总样本数**28056**， 其中正样本**2796**， 负样本**25260**。
- 随机取**5000**个样本训练， 其余测试。
- 样本归一化， 在训练样本上， 求出每个维度的均值和方差， 在训练和测试样本上同时归一化。

$$newX = \frac{X - mean(X)}{std(X)}$$

- 高斯核
- **5-fold cross validation**， 在
CScale = $[2^{(-5)}, 2^{15}]$; **gamma** = $[2^{(-15)}, 2^3]$;
上遍历求识别率的最大值。
上述**C**和**gamma**的区间设置参见**LIBSVM**自带的介绍：
a practical guide to support vector classification



用SVM解兵王问题

- 训练参数设置 **svmtrain(yTraining, xTraining, cmd)**
cmd参数如下:

(1) -s 0 "-s svm_type : set type of SVM (default 0)\n"
" 0 -- C-SVC (multi-class classification)\n"
" 1 -- nu-SVC (multi-class classification)\n"
" 2 -- one-class SVM\n"
" 3 -- epsilon-SVR (regression)\n"
" 4 -- nu-SVR (regression)\n"



用SVM解兵王问题

(2) -t 2

"-t kernel_type : set type of kernel function (default 2)\n"
" 0 -- linear: $u' * v$ \n"
" 1 -- polynomial: $(\gamma * u' * v + \text{coef0})^{\text{degree}}$ \n"
" 2 -- radial basis function: $\exp(-\gamma * |u - v|^2)$ \n"
" 3 -- sigmoid: $\tanh(\gamma * u' * v + \text{coef0})$ \n"
" 4 -- precomputed kernel (kernel values in training_instance_matrix)\n"



用SVM解兵王问题

(3) **-c CVALUE**

"-c cost : set the parameter C of C-SVC, epsilon-SVR, and nu-SVR (default 1)\n"

(4) **-g gammaValue**

"-g gamma : set gamma in kernel function (default 1/num_features)\n"

(5) **-v 5**

"-v n : n-fold cross validation mode\n"



浙江大学
ZheJiang University

用SVM解兵王问题

- 训练后获得的参数

(1) **$C = 16$, $\gamma = 0.0825$**

(2) 支持向量（即 α 不为0的向量）：**358个（162个正样本，196个负样本）**

(3) **$b = 6.2863$**



用SVM解兵王问题

■ 混淆矩阵

		预测	
实际		正样本	负样本
	正样本	True Positive (TP)	False Negative (FN)
	负样本	False Positive (FP)	True Negative (TN)

TP: 将正样本识别为正样本的数量（或概率）

FN: 将正样本识别为负样本的数量（或概率）

FP: 将负样本识别为正样本的数量（或概率）

TN: 将负样本识别为负样本的数量（或概率）



用SVM解兵王问题

■ 识别结果

		预测	
实际		正样本	负样本
	正样本	TP (507)	FN (5)
	负样本	FP (1)	TN (4487)

		预测	
实际		正样本	负样本
	正样本	TP (99.024%)	FN (0.976%)
	负样本	FP (0.022%)	TN (99.978%)

表1. 在训练集上的混淆矩阵
(Accuracy = 99.88%)



用SVM解兵王问题

■ 识别结果

		预测	
实际		正样本	负样本
	正样本	TP (2249)	FN (39)
	负样本	FP (51)	TN (20717)

		预测	
实际		正样本	负样本
	正样本	TP (98.295%)	FN (1.705%)
	负样本	FP (0.246%)	TN (99.754%)

表2. 在测试集上的混淆矩阵
(Accuracy = 99.60%)



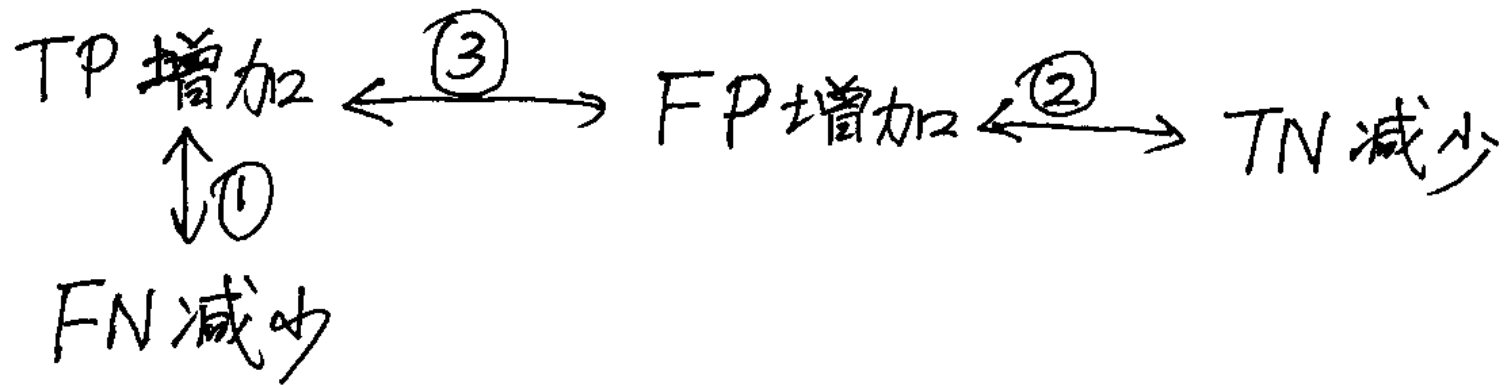
ROC曲线

■ 四个概率TP, FN, FP, TN的关系

1. $TP + FN = 1$

2. $FP + TN = 1$

3. 对同一个系统来说, 若TP增加, 则FP也增加。

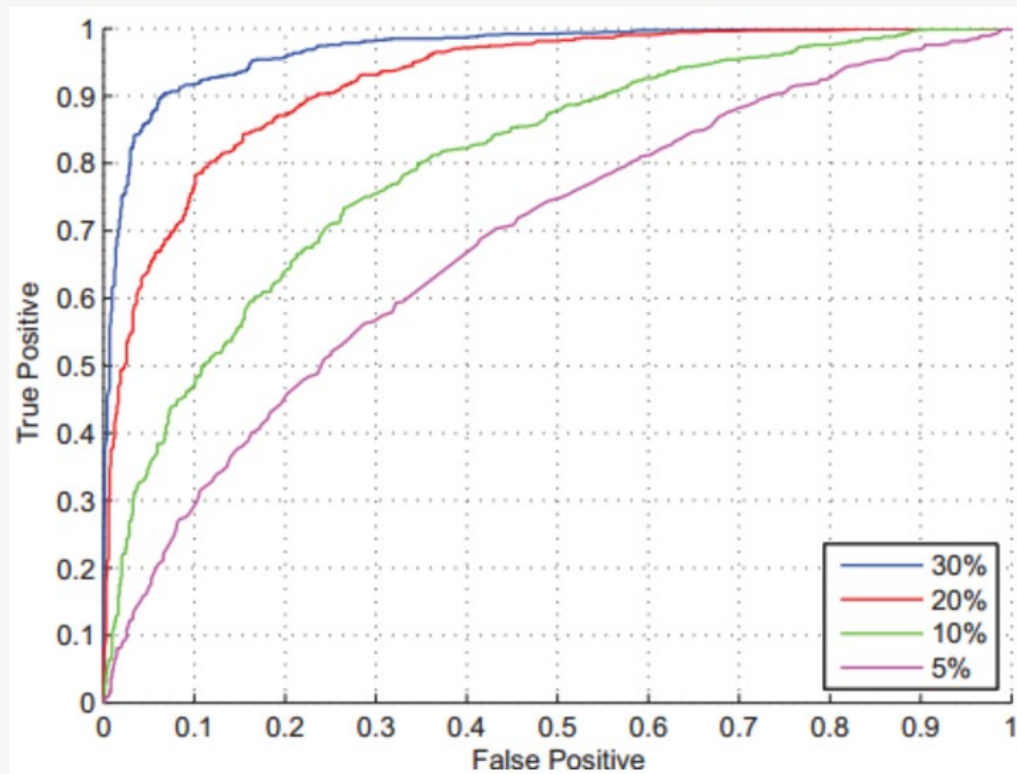




浙江大学

ZheJiang University

ROC曲线



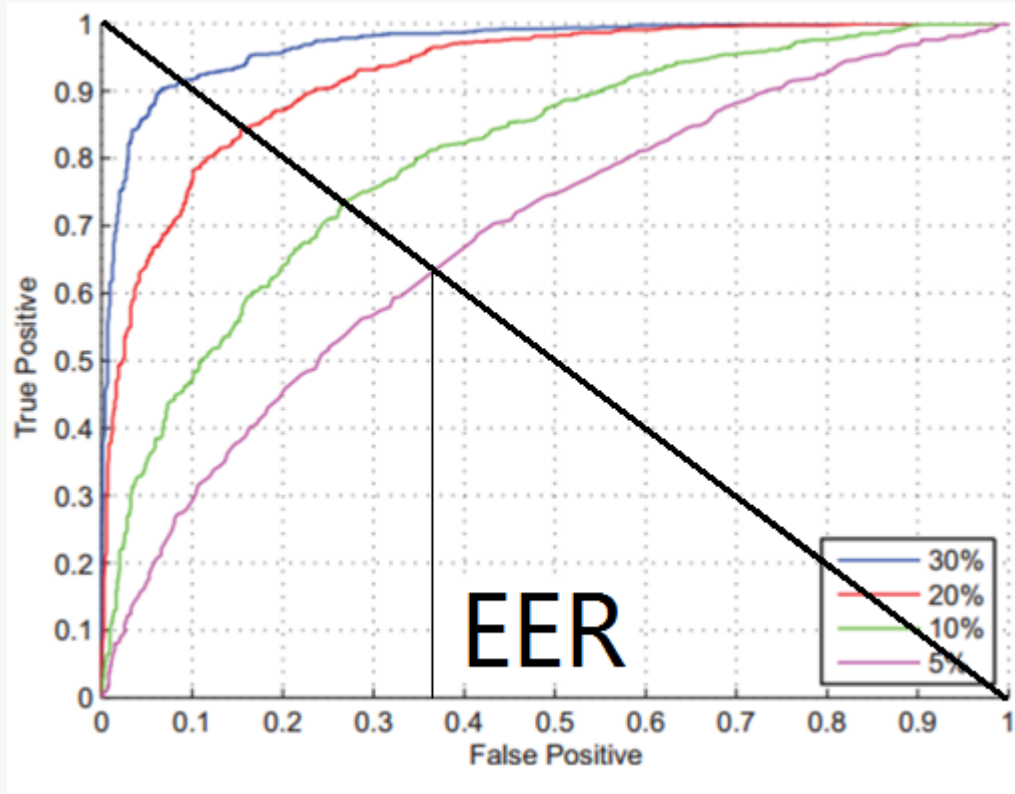
ROC (Receiver Operating Character)曲线，
是一条横坐标**FP**，纵坐标**TP**的曲线



浙江大学

ZheJiang University

ROC曲线



等错误率 (**Equal Error Rate, EER**)是两类错误**FP**和**FN**相等时候的错误率，可以直观的表示系统性能。



SVM处理多类问题

- **SVM**有三种方式处理多类问题，即类别大于**2**的问题：

(1) 改造优化的目标函数和限制条件，使之能处理多类
论文 **SVM-Multiclass Multi-class Support Vector Machine**

(2) 一类 **VS** 其他类

(3) 一类 **VS** 另一类