2024年1月12日 23:45

Final exam 5 questions, in 3Hrs

Homework week2, online, quiz at week3, 30-45mins.at most 1hr. homework hand written submission upload online

Nature inspired algorithm

P6/20 find the best input
8queens prob

P7 finite time slots

P9 xlsx for 8queens check

P12 Simulation for unknown new outputs

P16 目标函数给出了一个具体的数值评估方法
限制条件是二元评估

P18 Nondeterministic Polynomially，非确定性多项式

作业

2024年1月12日　23:47

Recap1

2024年1月12日　　23:45

BB/ Search Space/ Opt-CS/ Hardness

Motivation & inspiration
Framework
表现型和基因型，编码和解码的互相转换
Week2注重于突变等内容

2024年1月22日 星期一 18:49

P123 offspring another 347826591

学习P124 125的PMX 方法

Machine Learning

2024年1月12日　23:45

Mao kezhi

13周quiz两道题，1小时，15%分数

第五周不上课听录播

两个作业加起来15%，第六周一个，第八周一个，第十周一起交作业

期末1个遗传算法题，3个机器学习

contents

2024年2月5日 星期一 17:54

Data preparation
Supervised learning
Unsupervised Learning

Learn in three step

2024年2月5日 星期一　　18:45

Data input
Abstract
generalization

Machine Learning

1. Supervised learning:
   (1). Classification
   (2). Regression
2. Unsupervised Learning:
   (1). Clustering
   (2). Association analysis 不重要
3. Reinforcement Learning 不涉及

Data preparation

2024年1月12日 23:45

1. Qualitative data

Nominal(name, nationality, gender), can only count how many
ordinal(Grades ABC), can also compare, median quartiles, but not add, mean ,var

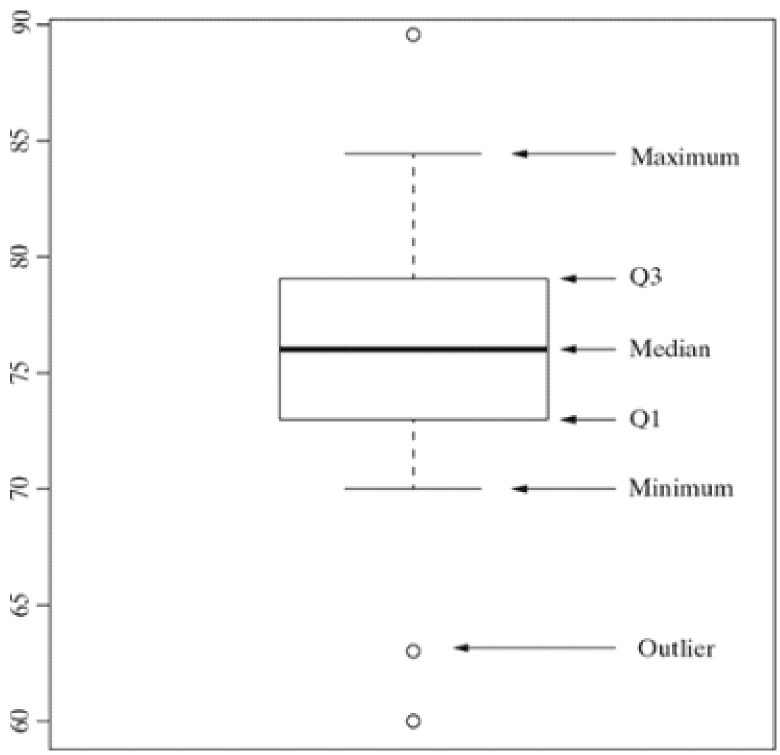
2. Quantitative data

Interval:(temperature) can add, subtract
ratio: all arithmetic operation

discrete or continuous

Box plot:

IQR: Q1 to Q3
The lower whisker extends up to 1.5 times of the inter-quartile range (or IQR) from the bottom of the box
The upper whisker extends up to 1.5 as times of the inter-quartile range (or IQR) from the top of the box



and Histogram

3. relationship

scatter plot, Two-way cross-tabulations

4. data remediation

outliers: mean and median differ a lot
Handling outliers
Remove/ imputation/ capping or build another model.

5. scaling/ normalization/ standardization

scaling, affect distance based algorithms, KNN SVM,
Gradient descent-based algorithms,

normalization: 0 to 1 , minmax scaling
standardization: (x-miu)/sigma

Bayesian rule

2024年1月12日 23:45

Posterior probability from: prior and evidence(product is joint prob)

Discriminator Func.

A Bayes classifier is determined by the conditional probability density function as well as the prior probability, equally posterior probability

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)p(\omega_i)$$

$$g_i(\mathbf{x}) = \log p(\mathbf{x}|\omega_i) + \log p(\omega_i)$$

the prior is obvious, but what is the conditional probability density function? we introduce [Gaussian PDF]

Gaussian distribution

2024年2月26日 13:30

1D gaussian
nD gaussain

$$p(\mathbf{x}|\omega) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

Need to estimate the mean vector and covariance matrix, so use the maximum likelihood estimation.

Estimated miu and sigma, quiet intuitive, right?

Case 2: Gaussian function with unknown $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$

With a similar deviation process, the maximum-likelihood estimate of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ are obtained as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$$

$$\hat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{k=1}^{n} (\mathbf{x}_k - \hat{\boldsymbol{\mu}})(\mathbf{x}_k - \hat{\boldsymbol{\mu}})^T$$

Gaussian mixture model

2024年2月26日 16:32

Step1: estimate γ from α, μ, Σ

Step2: update α, μ, Σ

Iteration stop criterion:

1. Loop count
2. Parameter change little

Assignment 1

2024年2月26日 星期一 18:33

Due 31.03.2024
Any lang

贝叶斯决策的理解

2024年3月23日　　16:45

进行分类做出决策：
去菜市场买菜，因为自己比较严谨，吃过所有的菜，根据过往多年经验知道这里有80%占比的菜好吃，20%不好吃，即为先验。与此同时默默记下好吃的菜大概是什么特征，也就是拥有了一些训练集。
有经验后买菜时候不会蒙眼随便拿去赌自己拿到好吃的菜，那样只有80%的几率，而要根据观测到的一些特征做出好不好吃的判断，发挥主观能动性。
特征比如有：颜色，大小，水分，表面光滑度。
某个特定的特征组合出现的概率即为evidence。因为以前有了prior，现在有了evidence，还需要一个好吃的菜里有这种特征的几率--likelihood，就能计算出这组特征的菜好吃的几率了。
怎么计算likelihood呢，这是一个条件概率：在"好吃"情况下，还有这种特征。
1.如果以前经常出现这种特征，已经足以达到统计学标准，直接根据以往频次计算频率得到概率。
2.但如果以前好吃的里面没有完全匹配的情况怎么办？这很常见，毕竟世界上没有一模一样的菜。只能估计，比如（深绿，5cm，富含水分，很光滑）这组特征是好吃的几率很高，现在的（绿，6cm，富含水分，很光滑）好吃的几率肯定也不低，具体数值怎么算呢，只能假设这些多维特征符合多维的高斯分布。但有可能不止1个高斯分布成分，而是多个。
只要有均值向量，协方差矩阵，就能描述出来一整个高斯分布，但贝叶斯决策要解决的是无穷无尽数据的情形，永远不可能知道参数的真实值。所以需要估计这两个参数，或者n*2个参数。
估计过程使用极大似然估计，也就是说，根据我们的训练样本，找到某个特定的参数向量θ，使得训练数据符合该特定高斯分布的概率达到了极大值，因为根据我们直觉，既然我们抽样到这些样本，那我们理所当然可以认为正是因为这种结果出现概率最大，所以我们才得到这个结果。
使用带分类的数据训练，我们假设就按1个高斯来算，求极值要算导数或者偏导等于0的点，不必考虑严格证明，复杂运算后结果很符合直觉的是，高斯估计均值就是所有样本均值，高斯估计协方差就是样本协方差。训练过程就是计算这2个参数的过程，数据越多，参数越准确，然后有了它，就是有个pdf（概率密度函数）。
但是计算多高斯时候并不能直接这么算，而是要用EM算法迭代求解。
接下来就能计算刚才想要的似然了，输入当前特征，得到似然，因为我们计算的都是"好吃的"pdf，所以似然就是条件概率。
再然后就得到后验概率，如果概率大于50%，那我们就认为该菜是好吃的了。

naïve bayes

2024年4月26日 15:29

Then the posterior probability is:

$$p(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)p(\omega_j)}{p(\mathbf{x})} = \frac{\prod_{i=1}^{n} p(x_i|\omega_j)p(\omega_j)}{p(\mathbf{x})}$$

Gaussian Naive Bayes, for each dim we have

$$p(x|\omega_j) = \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right)$$

Bernoulli Naive Bayes

don't need pdf, just count independent features conditional probability

Multinomial Naïve Bayes

To address the issue, we can use the so-called Laplace smoothing (add 1):

$$p(x_i|\omega_j) = \frac{counts(x_i, \omega_j) + 1}{\sum_{k=1}^{d}(counts(x_k, \omega_j) + 1)}$$

$$= \frac{counts(x_i,\omega_j)+1}{(\sum_{k=1}^{d} counts(x_k,\omega_j))+d}$$

☐ is the number of unique words in all training data

LDA

2024年2月26日 17:17

To simplify the classifier, use a line rather than a curve.
It is a 2 class question.

Linear discriminant function/ or hyperplane

The func g(x)

$$g(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$$

$$r = \frac{g(\mathbf{x})}{||\mathbf{w}||}$$

Is related to distance r(to the hyperplane)

in page 11, we try to find the g(x):
we hope intraclass projection more concentrated,
while interclass more separated.

The Fisher linear discriminant attempts to find such **w** that the following criterion is maximized:

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

so we have:

Scatter matrix

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

$$\boldsymbol{S}_W = \boldsymbol{S}_1 + \boldsymbol{S}_2$$

W denotes withinclass
B denotes betweenclass

Thus, the maximization criterion $J(\mathbf{w})$ can be expressed as:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

17

the final result is:

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

and

$$\boldsymbol{S}_W = \boldsymbol{S}_1 + \boldsymbol{S}_2$$

15

$$\mathbf{w} = \mathbf{S}_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

If matrix $\mathbf{S}_W$ is singular, i.e.

$$|\mathbf{S}_W| \approx 0$$

Then we can use the so-called regularization technique:

$$\mathbf{w} = (\mathbf{S}_W + \beta \mathbf{I})^{-1}\,(\mathbf{m}_1 - \mathbf{m}_2)$$

$$w_0 = -\frac{\mathbf{w}^T(\mathbf{m}_1 + \mathbf{m}_2)}{2}$$

MDA in 36

2024年2月26日 星期一 20:58

Pca dimensional reduction is unsupervised
Here we need reduction supervised

multiple classification need many discrimination function gi(x).

SVM

2024年1月12日 23:45

2class with label +1 -1
Optimal hyperplane when margin of separation is maximized.

对于任何一个输入的向量x，计算判别函数g（X）后除以W的模得到距离判别超平面的距离r

$$r = \frac{g(\mathbf{x})}{||\mathbf{w}||}$$

支持向量是指距离正负1决策边界最近的数据点（向量）。也就是支持向量是最难分类的点，最重要的点

此时为一个限制条件的最优化问题，限制条件为

$$d(i) \times [\mathbf{w}^T \mathbf{x}(i) + b] \geq 1$$

$$i = 1,2, \dots N$$

Where $d(i)$ is the class label of sample $\mathbf{x}(i)$. It takes the value of $+1$ or -1 for class 1 and 2, respectively.

优化W的2范数为最小

$$\rho = \frac{2}{||\mathbf{w}||}$$

that they satisfy the constraints

$$d(i) \times [\mathbf{w}^T \mathbf{x}(i) \ + b] \geq 1 \quad \text{for } i = 1,2, \dots, N$$

that they satisfy the constraints

$$d(i) \times [\mathbf{w}^T \mathbf{x}(i) \ + b] \geq 1 \qquad \text{for } i = 1,2, \dots, N$$

and the weight vector minimizes the following cost function:

$$J(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T \mathbf{w}$$

根据拉格朗日乘数法得到的损失函数L(w,b,a)函数的偏w导等于0我们有

$$\mathbf{w} = \sum_{i=1}^{N} \alpha(i) d(i) \mathbf{x}(i)$$

相较于上节课的LDA计算特征向量，SVM只需要计算这些输入的线性和。

对于b的偏导得0有

$$\sum_{i=1}^{N} \alpha(i) d(i) = 0$$

+1分类的所有a(i)和等于所有-1分类的。

根据KKT理论，就能得知那些支持向量的a(i)才能大于0，不是支持向量的，a(i)等于0.

最后在p21得到一个不用w表示的损失函数L，由此可以替换为对偶问题

工具计算出最优化w

$$\mathbf{w}^* = \sum_{i=1}^{N} \alpha(i) d(i) \mathbf{x}(i)$$

由此得到b

$$b^* = \frac{1}{d(i)} - (\mathbf{w}^*)^T \mathbf{x}(i) = d(i) - (\mathbf{w}^*)^T \mathbf{x}(i)$$

$$b^* = \frac{1}{N_s} \sum_{\alpha(i) \neq 0} [d(i) - (\mathbf{w}^*)^T \mathbf{x}(i)]$$

P24，根据算出来优化的W，然后根据一个输入x就能得到b，可以算他们的平均来确定b

P32， 很多情况下这两类数据并不是线性可分的。

Classification trees

2024年1月12日 23:45

Euclidean distance used for measure similarity

A nonmetric method, classification tree

Pros: interpretability
Incorporate prior knowledge

分类树的节点问题选择：大道致简

定义杂质impurity指标：

Entropy impurity

$$i_E(N) = -\sum_j P(\omega_j)\log_2 P(\omega_j)$$

Variance and gini impurity

Var for 2class gini for multiclass(actually covariance matrix)

$$i_{Var}(N) = P(\omega_1)P(\omega_2)$$

ralization of the variance impurity, applicable to
e category, is the *Gini impurity*:

$$i_{Gini}(N) = \sum_{j\neq k} P(\omega_j)P(\omega_k) = \frac{1}{2}\left[1 - \sum_j P^2(\omega_j)\right]$$

Misclassification impurity

$$i_{MC}(N) = 1 - \max_j P(\omega_j)$$

when splitting, branch impurity is weighted gini and sum. choose the lowest splitting method.

stop iteration threshold, trade off

$$J = \alpha \times size + \sum_{N \in leaf\ node} i(N)$$

typical trees:
CART/ ID3/ C4.5/ Random forest

random forest:
n subsets, use of each subsets m features less than original d features, sample with replacement
when predicting, use all trees, vote for result.

Metrics

2024年1月12日 23:45

19:50讲了一句话

Regression

2024年3月25日 星期一 20:15

Feature selection

2024年1月12日 23:45

Last topic on supervised learning

Curse of dimensionality: 1. data sparsity 2. distance concentration

Select subset of features: 1. better generalization

Feature selection: just subset of original features
Feature extraction: create new feature hyperspace

Peaking Phenomenon: there is a critical point to get best performance in selecting L features.

2 method in individually feature evaluation: only to find bad features.
Based on separability: Fisher ratio, for continuous variables
Based on relevance: Mutual information, use x entropy and y entropy and joint entropy, for discrete features.

fisher

$$R = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}$$

MI:

$$MI(x,y) = E(x) + E(y) - E(x,y)$$

Feature subset selection:

1. Search algorithm
2. Evaluation criterion

Search: 1. exhaustive search 2.sequential forward selection 3.sequential backward elimination

Evaluation criterion for feature subsets:

1. Use performance after training and test
2. Use separability measures, not related to classifiers, mahalanlbis measure/ scatter measure/

mahalanibis

$$J_{1,2} = (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{C}^{-1} (\mathbf{m}_1 - \mathbf{m}_2)$$

c is cov

scatter

$$J = \mathrm{Tr}(\mathbf{S}_W^{-1} \mathbf{S}_B)$$

$$J = \frac{\mathrm{Tr}(\mathbf{S}_B)}{\mathrm{Tr}(\mathbf{S}_W)}$$

3 methods in subset selection:

Filter method: use separability measure and no classifier in loop
Wrapper method: use classifier
Embedded method: lasso regression use L1 norm shrinks the coefficients associated with the least important features to zero. Also need to train a model to predict label y_hat. Many models.

10.Clustering

2024年1月12日 23:45

Types: centroids-based, hierarchical, distribution-based, density-based

Centroids-based:

K-means:

WCSS define total squared distance(error)
WCSS can be used as criterian as long WCSS minimized.
WCSS suitable when samples seperated well.

$$WCSS = \sum_{i=1}^{k} \sum_{\mathbf{x} \in D_i} \|\mathbf{x} - \mathbf{m}_i\|^2$$

Iteration:

1. Set K
2. Random K centroids
3. Traverse all samples assign to nearest centroids.
4. Calc new centroids.
5. Repeat 3-4 until no change to centroids.
6. Try other K, repeat 1-5 use Elbow Method to decide best K

Hierarchical clustering:

Agglomerative: bottom-up

1. Each sample as a cluster
2. Compute Euclidean distance
3. Merge 2 nearest samples
4. Compute Euclidean distance and single linkage (nearest neighbour) or complete linkage or centroid linkage or average linkage

Divisve: top-down

Distribution-based clustering:

GMM model:

1. Set m clusters
2. Init a,u,y
3. EM
4. Mean of GMM is cluster centres.
5. Gmm = fitgmdst(data, m);

Density-based clustering:

DBSCAN:

minPts, can be choosed at least 3, D+1 , better 2D
Eps , should small, elbow

Step1: set minPts, eps, random start point
Step2: find density connected points
Step3: Choose another start point
Step4: finish when no further change

Pros: reduce noise, no K hyperparameter, arbitrary shape
Cons: rely density, rely init, hard esp

11.clustering evaluation

2024年4月8日 星期一　　20:43

silhouette coefficient

$$S_i = \frac{d_{bi} - d_{si}}{max\{s_{bi}, d_{si}\}}$$

$$S_{overall} = \frac{1}{N}\sum_{i=1}^{N} S_i$$

Dunn Index:

$$DI = \frac{\min\limits_{k,l\in m} d_{kl}}{\max\limits_{k\in m} \delta_k}$$

Davies-Bouldin index:

Calinski-Harabasz Index:

Rand Index:

Use for vote, for compare 2 clustering method.

quiz

2024年1月12日 23:45

1. kfold
2. cluster

Q&A

2024年4月8日 星期一 20:43

Expectation-Maximization:
After searching some related papers, I found most of the implementations use a diagonal matrix as initial covariance matrix, but I tried some random positive definite matrices, and still have an acceptable convergence speed. When I try the matrix in lecture note lecture note 3.4.2, sigma_hat_21, I can't get a result, I think this mat has a negative determination if I'm not mistaken. Could you please check that? And what's the best choice of initial value?

review

2024年4月15日 4:05

7.Performance Evaluation

hold out method, use 75% training, 25% testing
or 70% training, 15% validation, 15% testing
imbalance problem, use stratified random sampling

K fold: LOOCV, one sample/fold
repeated K fold

metric:

Acc	correct ratio
Sensitivity	in positive, correct ratio
Specificity	in negative, correct ratio
Precision	in prediction, correct p ratio
recall	in prediction, find p ratio

F score= 2pr/p+r

ROC: even having a model, dealing with some problem still need to add other scaling, so we can find a best result
ROC is tpr/fpr
AUC

8.Regression

least square estimation:

We have:

$$\mathbf{y} = \mathbf{\Phi} \times \mathbf{\theta} + \mathbf{\varepsilon}$$

Define the following loss function:

$$J = \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{2}\mathbf{\varepsilon}^T \mathbf{\varepsilon}$$

find the best theta

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\mathbf{y}$$

L2-norm: ridge estimation

$$\hat{\boldsymbol{\theta}}_{ridge} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi} + \lambda\mathbf{I})^{-1}\boldsymbol{\Phi}^T\mathbf{y}$$

Where $\mathbf{I}$ is a $(m+1) \times (m+1)$ identity matrix.

L1-norm: Lasso, produce sparse model, eliminating unimportant var coefficient
large lambda make sparse model.

regression metric:

MSE, RMSE, MAE, R-squared(1-RSS/TSS) residual sum square total sum square

$$Adjusted\ R^2 = 1 - \frac{(1-R^2)(n-1)}{(n-m-1)}$$

validation:

linear, NORMAL error, error mean 0, multivariate normal, Homoscedasticity

9.Feature Selection

Feature selection

10.Clustering

10.Clustering

11.Cluster Evaluation

11.clustering evaluation

Genetic Algorithms

2024年4月23日 21:48

1. Evolutionary computing

4 problems:

black box:
when unknown inputs, black box optimization question, timetable/ TSP/ satellite
when unknown model, modelling problems, can be transformed into optimisation problems.
when unknown outputs, simulation,

search problems: simulation// optimization modeling(search space)

optimization constraint satisfaction

	Objective function	
Constraints	Yes	No
Yes	Constrained optimisation problem	Constraint satisfaction problem
No	Free optimisation problem	No problem

NP problems: P NP NPC NPH
EA: reproduction/ mutation

2. EA components:

representation, evaluation(fitness function), population, selection mechanism, variation operators(mutation crossover), initialization termination,

GA is the algorithm for binary string.

8Queens problem: phenotype and genotype

simple GA:

1	2	3	4	5	6	7	8	9	10	11	12
String No.	Initial Popn.	x	$f(x)$	% of Total	No. Sel.	Mating Pool	Mate	C'over Site	New Popn.	x	$f(x)$
1	01101	13	169	14.4	1	01101	2	4	01100	12	144
2	11000	24	576	49.2	2	11000	1	4	11001	25	625
3	01000	8	64	5.5	0	11000	4	2	11011	27	729
4	10011	19	361	30.9	1	10011	3	2	10000	16	256
Sum			1170	100.0	4						1754
Average			293								439

representation: float to L binary, in order to perform crossover and mutation.

Ackley is used to test how good is your GA program.

$$f(x) = -20 \cdot \exp\left(-0.2\sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}\right)$$
$$-\exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$$

$$\Gamma(a_1, \dots, a_L) = x + \frac{y-x}{2^L - 1} \cdot \left(\sum_{j=0}^{L-1} a_{L-j} \cdot 2^j\right) \in [x, y]$$

mutation: uniform mutation, self-adaptive mutation

crossover: Single arithmetic crossover, Simple arithmetic crossover, Whole arithmetic crossover, blend crossover,

multiparent recombination:
n-1 point, diagonal.
arithmetic crossover

representation: permutation,
swap mutation, insert mutation, Scramble mutation, Inversion mutation,
order 1 crossover: 1. substring copy to offspring, permutate other numbers in origin order, 2 parents 2 children
PMX crossover:
Cycle crossover:
edge recombination:

Element	Edges	Element	Edges
1	2,5,4,9	6	2,5+,7
2	1,3,6,8	7	3,6,8+
3	2,4,7,9	8	2,7+, 9
4	1,3,5,9	9	1,3,4,8
5	1,4,6+		

Choices	Element selected	Reason	Partial result
All	1	Random	[1]
2,5,4,9	5	Shortest list	[1 5]
4,6	6	Common edge	[1 5 6]
2,7	2	Random choice (both have two items in list)	[1 5 6 2]
3,8	8	Shortest list	[1 5 6 2 8]

Element	Edges	Element	Edges
1	2,5,4,9	6	2,5+,7
2	1,3,6,8	7	3,6,8+
3	2,4,7,9	8	2,7+, 9
4	1,3,5,9	9	1,3,4,8
5	1,4,6+		

Choices	Element selected	Reason	Partial result
All	1	Random	[1]
2,5,4,9	5	Shortest list	[1 5]
4,6	6	Common edge	[1 5 6]
2,7	2	Random choice (both have two items in list)	[1 5 6 2]
3,8	8	Shortest list	[1 5 6 2 8]
7,9	7	Common edge	[1 5 6 2 8 7]
3	3	Only item in list	[1 5 6 2 8 7 3]
4,9	9	Random choice	[1 5 6 2 8 7 3 9]
4	4	Last element	[1 5 6 2 8 7 3 9 4]

representation: tree

Machine Learning

2024年4月23日 21:49

1. Introduction
   Learn in three step
2. Data preparation
   Data preparation
3. Bayesian
   课时6
4. LDA
   LDA
5. SVM
   SVM
6. Decision Tree
   Classification trees
7. Feature Selection
   Feature selection
8. Evaluation
9. Regression
10. Clustering
11. Clustering Evaluation
    课时13

Revision2024