

Decision Tree

- ↳ minimizing standard deviation \leftarrow continuous data
- ↳ famous of overfitting.

Cross validation

~~gini~~ Decision Tree Classifier

Criterion:

- gini \rightarrow Gini impurity
- $-\log_{10} \{ \}$ Shannon Information Gain
- entropy

Splitter:

- best
- random

Warum benötzen? \leftarrow erklären / Begründen

- ↳ Impul...
↳ Ordinal ...
↳ scalar ...
↳ Algorithmo ...
- } Prozess ...?

SVM = Support Vector Machine

- ↳ takentime in sklearn.

```
from sklearn.datasets import make_blobs
```

← New
version

Steigungsfaktor = gradient

Achsenabschnitt = fit.b.intercept_

Randbereich = margin

SVM = Support Vector Machine

↳ Standard Scaler \rightarrow for SVM \leftarrow so scaling may be needed

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$$

$$\gamma = \frac{1}{2\sigma^2} \quad \Leftarrow \text{for GridSearch}$$

$$k(x, x') = \exp\left(-\gamma \|x - x'\|^2\right)$$

\uparrow
Gamma in GridSearch

γ = gamma in SVC in sklearn

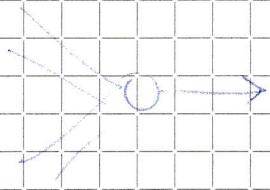
gs = GridSearchCV (svc, parameter)

gs.fit (X, y)

best = gs.best_estimator_

best.predict (Xnew)

Neuronale Netzwerke: \rightarrow Kapitel 7 Früchte



$$e = \hat{y} - y$$

def: residual
outflow

$$\hat{y}_i = \sum_j w_{ij} x_j$$

$$\frac{\partial e}{\partial y_i} = -1$$

$$\frac{\partial e}{\partial w_j} = \cancel{\sum_i} \cancel{\hat{y}_i} x_j$$

$$\frac{\partial e}{\partial w} = x_j$$

$$\frac{\partial e^2}{\partial w} \rightarrow \frac{\partial e}{\partial w} \cdot 2e = 2e x_j$$

Projekt

With score many

- CV - score \leftarrow cross validation

- gs - score \leftarrow grid search score

def __init__(self, scalar=True, scaler=MinMaxScaler()):

 self.sc = scaler

 self.scalar = scalar

def fit(self, X, y)

 if ~~scalar~~ self.scalar:

 X_shal = self.sc.fit_transform(X)

 else

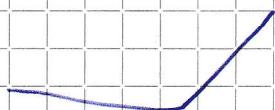
 X_shal = X

predict(\tilde{X})

\hookrightarrow scale or not scale...

Activation function

RELU



Batch Learning & Incremental Learning

Voting Classifier \rightarrow for project ???

\hookrightarrow find best Classifier.

} vs. Grid Search

Random Forest

↳ has many trees

Bagging:

↳ Bagging Classifier & Bagging Regressor.

Bootstrap ↲ in Random Forest

bob-score ↲ out-of-bag ↳ only if Bootstrap is true

↳ RandomForestClassifier

Bagging vs Boosting

Dimensionality Reduction } Feature Engineering
 - feature selection
 - PCA

Composite Estimator

PCA (Hauptkomponentenanalyse)

- Kovarianz-Matrix

Zentriert: $(\mathbf{x} - \bar{\mathbf{x}})$ ← mean will be zero

$\mathbf{x} = \mathbf{x} - \mathbf{x}_{\text{mean}}$

np.linalg.eigh (M)

↓
eigenvekt.

eigh → eigen von hermitian
 eig → eigen

Hermitian Matrix \Rightarrow symmetrisch, diagonale Spiegel
 diagonal mirror

pca. components_ ← ~~all~~ eigen vectors

pca. singular-values ← eigen values

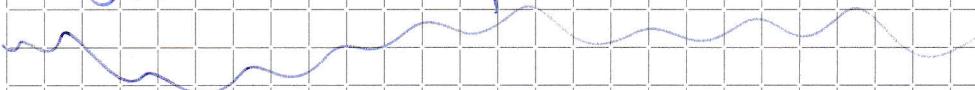
← sorted from large to small

pca. explained-variance ← sorting pca. components based on it

pca. noise-variance

pca. explained-variance-ratio.

TSNE ~ Manifold



Projekt:

↳ Ziel von Projekt ① Richtigkeit
↓ was ist minimiert/maximiert

Recall = $\frac{\text{gefundene in Vorrau}}{\text{alle ech.}}$

Precision = $\frac{\text{richtig in Vorrau}}{\text{alle Vorburau}}$

Bericht → PDF
↳ Wind

Freitag, 27.10.2023

Clustering

↳ DB SCAN

↳ Hierarchical Clustering

↳ K-Means

DB SCAN ?

↳ Kernpunkt

↳ Randpunkt

↳ Rauschpunkt

Kann es nicht random

when is cluster good?

KD-Baume ..