scikit-learn & Bokeh

David Dobrovolný

scikit-learn

- `https://scikit-learn.org`
- built on NumPy, SciPy and matplotlib
- open source
- on-going development (6 updates in 2020, as of May 4th 2021, 2 updates in 2021)
- used by e.g. JPMorgan, Spotify, Télécom ParisTech, Booking.com

scikit-learn

- classification
- regression
- clustering
- dimensionality reduction
- model selection
- preprocessiong

Naive Bayes

```
gnb = GaussianNB()
model = gnb.fit(X_train, y_train)
model.predict(X_test)
```

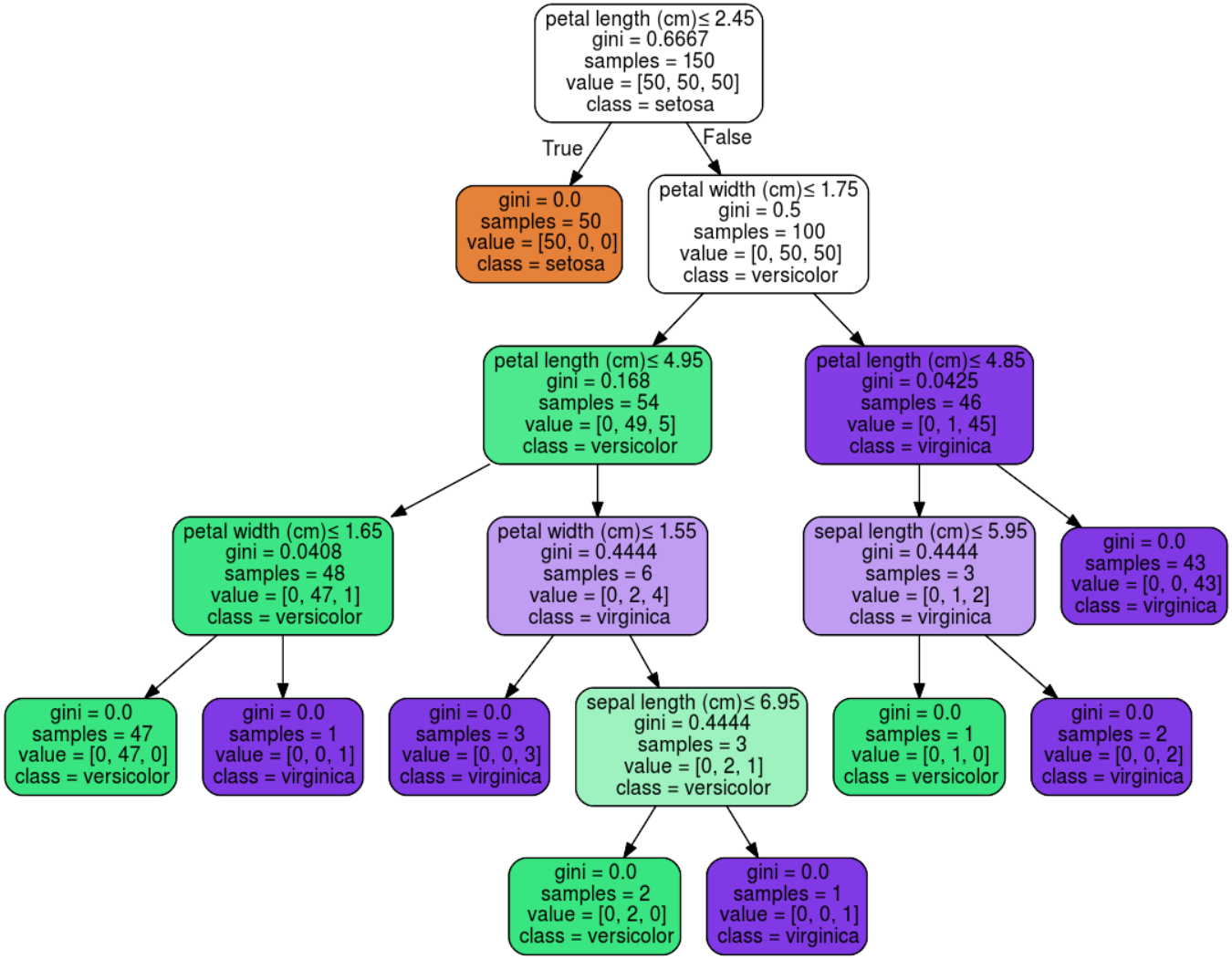
- Gaussian
- Multinomial
- Bernoulli
- ...

Decision Trees

```
clf = tree.DecisionTreeClassifier()
clf = clf.fit(X, Y)
clf.predict(test)
clf.predict_proba(test)
```

- ID3, C4.5, C5.0 and CART
- classification and regression
- Maximum depth, minimum leaf samples, impurity decrease, ...
- Can export to Graphviz format (next slide)

Decision Trees



Support Vector Machines

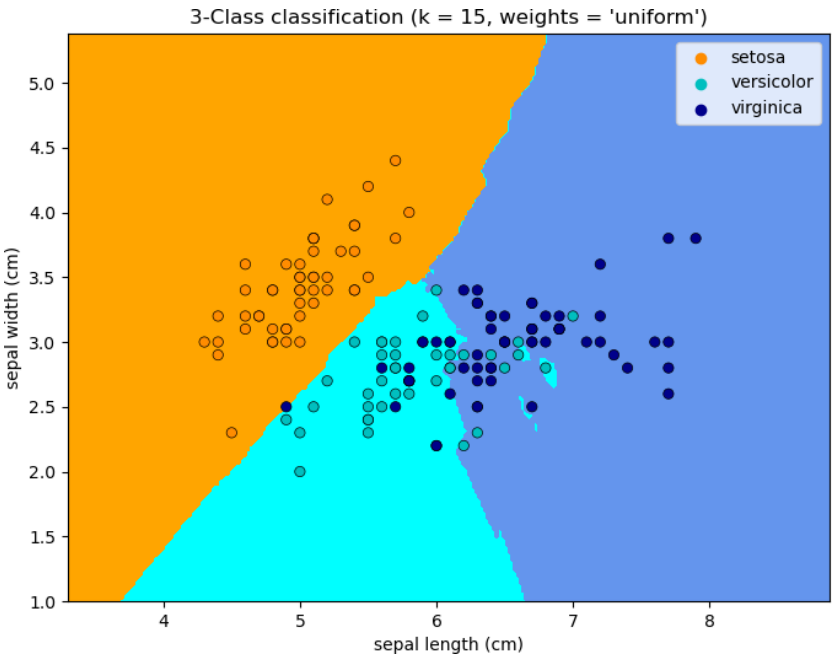
```
clf = svm.SVC()
clf.fit(X, y)
clf.predict(test)
```

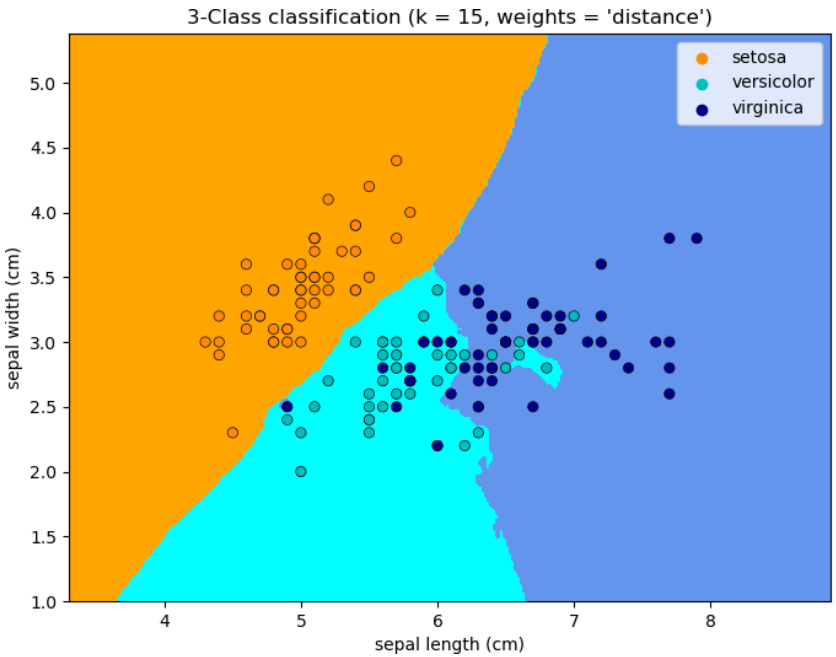
- classification and regression
- Different kernel functions
  - Linear: $\langle x, x' \rangle$
  - Polynomial: $(\gamma \langle x, x' \rangle + r)^d$
  - RBF: $\exp(-\gamma \parallel x - x' \parallel^2)$
  - Sigmoid: $\tanh(\gamma \langle x, x' \rangle + r)$
  - custom

Nearest Neighbours

- unsupervised
  - BallTree
  - KDTree
  - brute force
  - auto (algorithm determines the best approach)
- classification
  - uniform weights
  - weights based on distance (next slide)
- regression

Nearest Neighbours

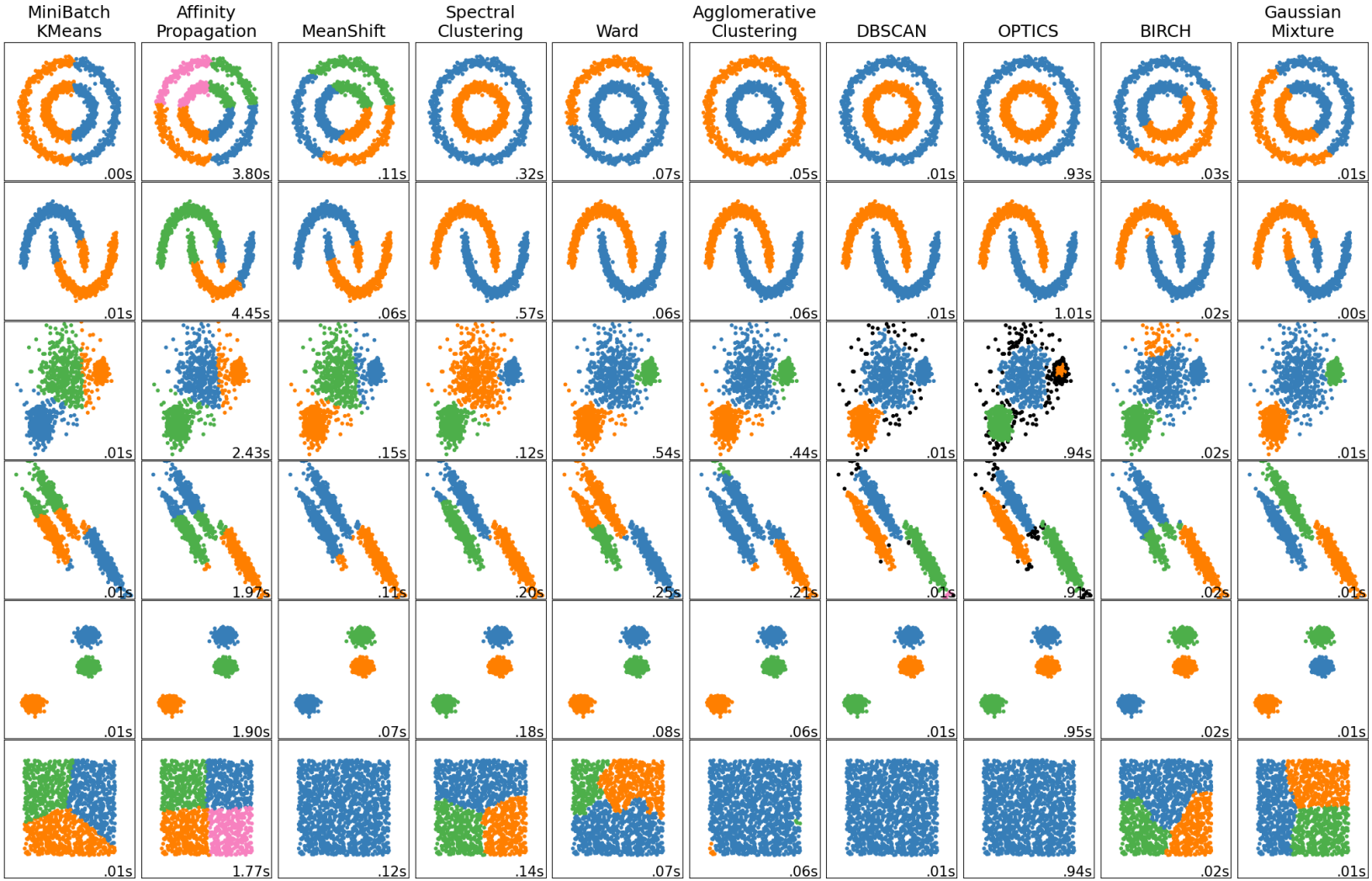




Other

- Clustering (next slide)
- Ensemble methods
  - Random Forests
  - Adaboost
- Semi-supervised learning
  - Self Training
- Neural Networks

Clustering


MiniBatch KMeans
Affinity Propagation
MeanShift
Spectral Clustering
Ward
Agglomerative Clustering
DBSCAN
OPTICS
BIRCH
Gaussian Mixture
.00s
3.80s
.11s
.32s
.07s
.05s
.01s
.93s
.03s
.01s
.01s
4.45s
.06s
.57s
.06s
.06s
.01s
1.01s
.02s
.00s
.01s
2.43s
.15s
.12s
.54s
.44s
.01s
.94s
.02s
.01s
.01s
1.97s
.11s
.20s
.25s
.21s
.01s
.91s
.02s
.01s
.01s
1.90s
.07s
.18s
.08s
.06s
.01s
.95s
.02s
.01s
.01s
1.77s
.12s
.14s
.07s
.06s
.01s
.94s
.02s
.01s

Bokeh

- `https://bokeh.org`
- Interactive web browser visualizations.
  - Server App - allows more interactive manipulation
  - Notebook
  - Standalone - limited interactivity, produces html file
  - Examples: `https://docs.bokeh.org/en/latest/docs/gallery.html`