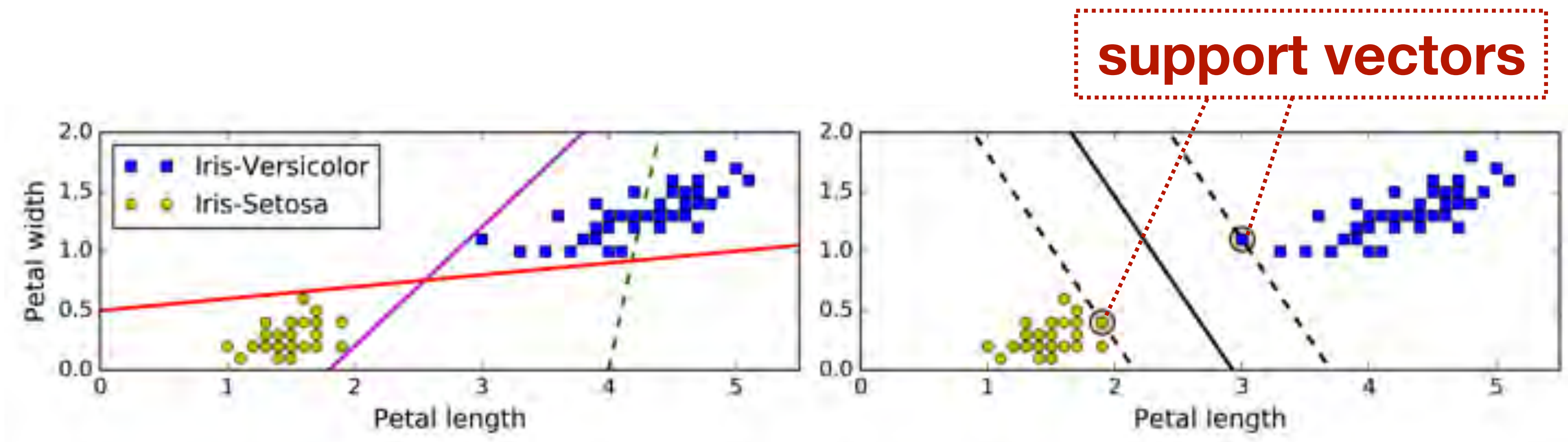
Advanced Data Analysis

DATA 71200

Class 10: Support Vector Machines and
Uncertainty Estimates from Classifiers

Linear Support Vector Classifier

- **Find the linear classifier with the best separation (margin) between the two classes**
    - Operationally the data points used to make this calculation are the support vectors



Three possible linear classifiers

Linear classifier with the widest margin

Géron, Aurélien. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras and TensorFlow* O'Reilly Media, Inc.

Support Vector Machines

- **Linear models are limited to separating classes with lines (hyperplanes in higher dimensions)**

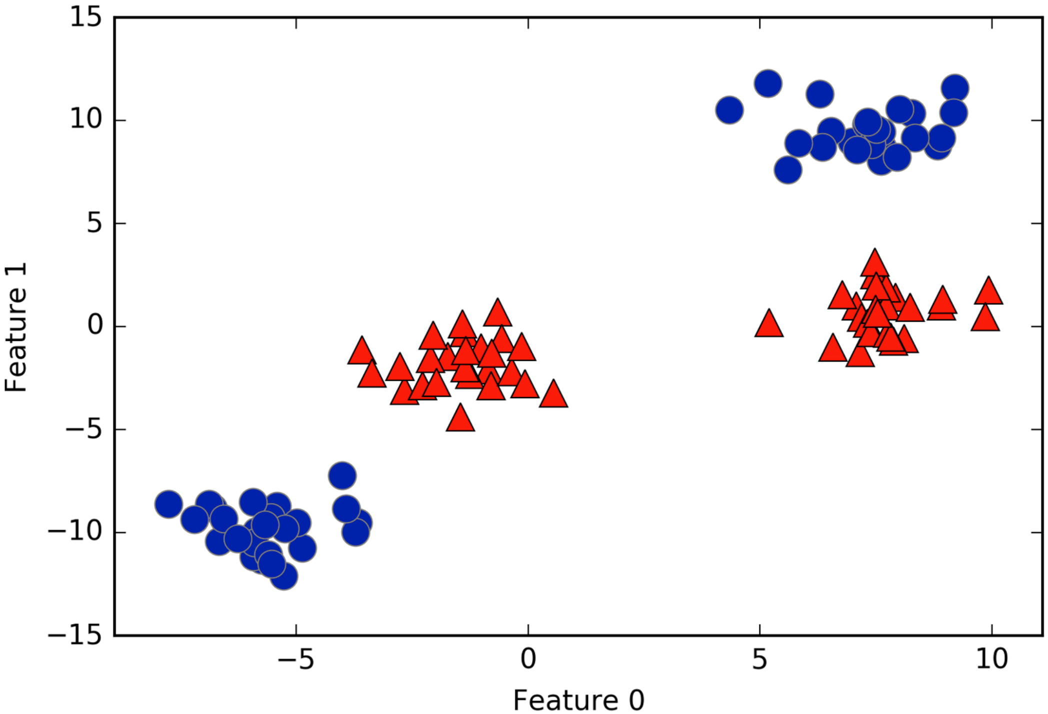


Figure 2-36. Two-class classification dataset in which classes are not linearly separable

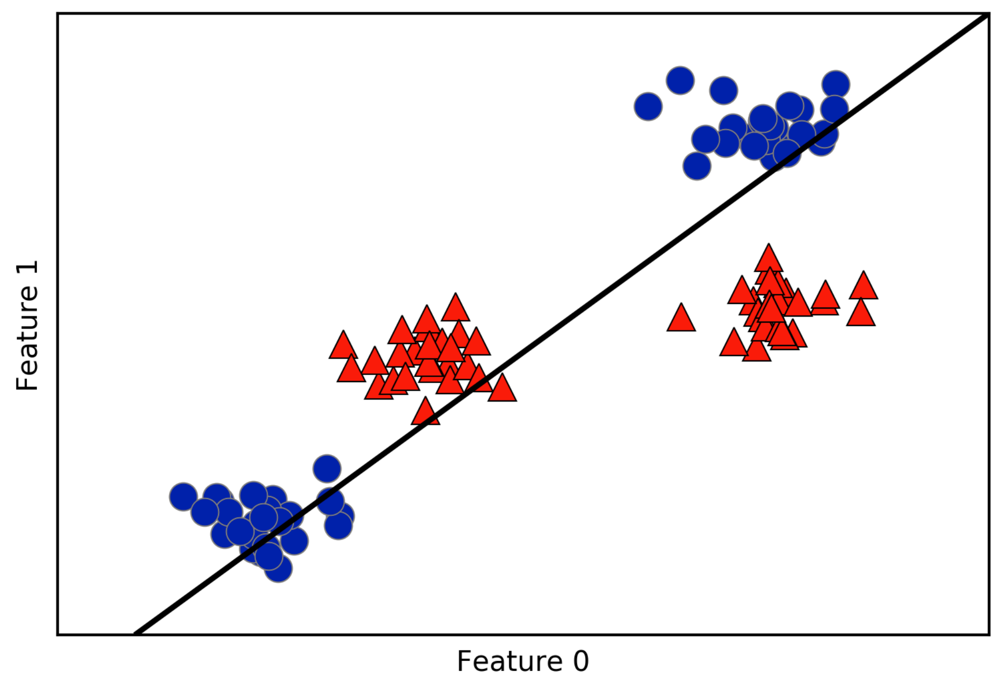


Figure 2-37. Decision boundary found by a linear SVM

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.

Support Vector Machines

- **Data can be transformed into a higher dimension**
- **Where an effective linear hyperplane can be fit to separate the data classes**

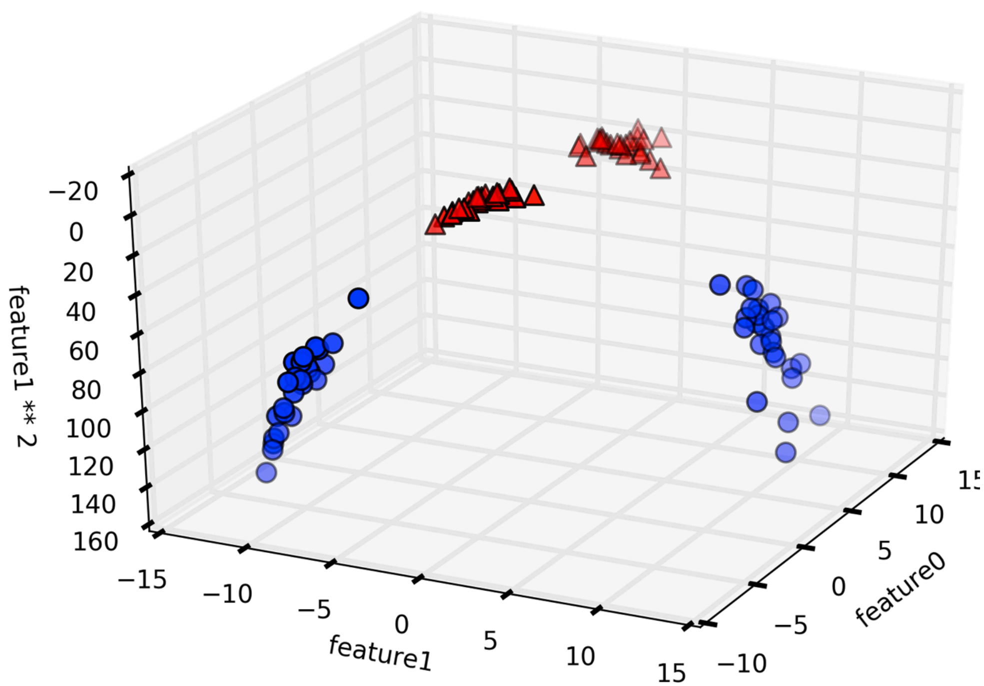


Figure 2-38. Expansion of the dataset shown in Figure 2-37, created by adding a third feature derived from feature1

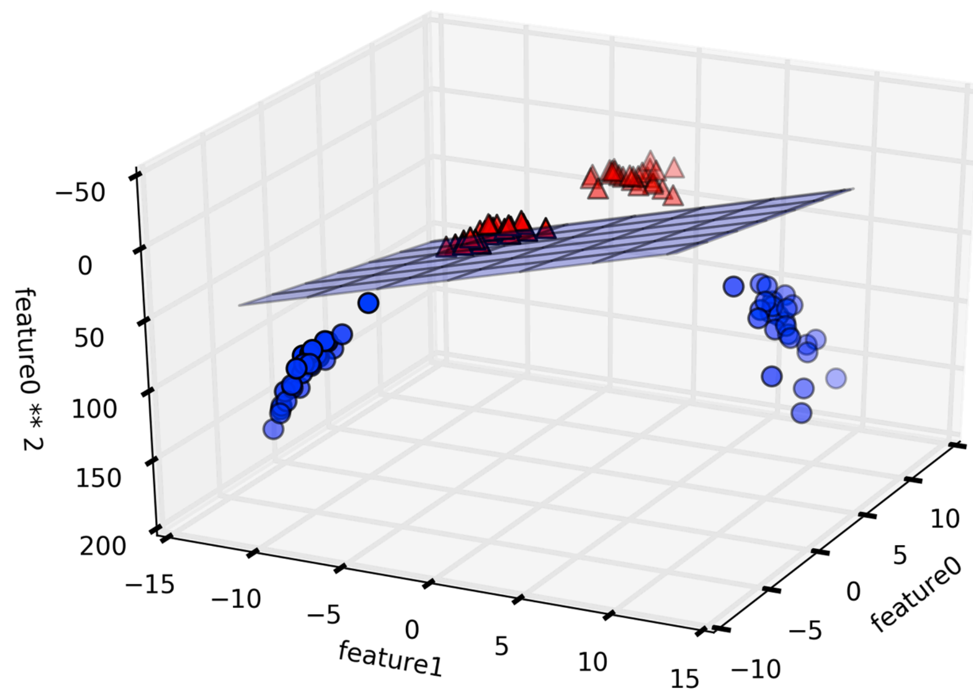


Figure 2-39. Decision boundary found by a linear SVM on the expanded three-dimensional dataset

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.

Support Vector Machines

- **When projected back to the two original features (2D) the resultant decision boundary is not linear**

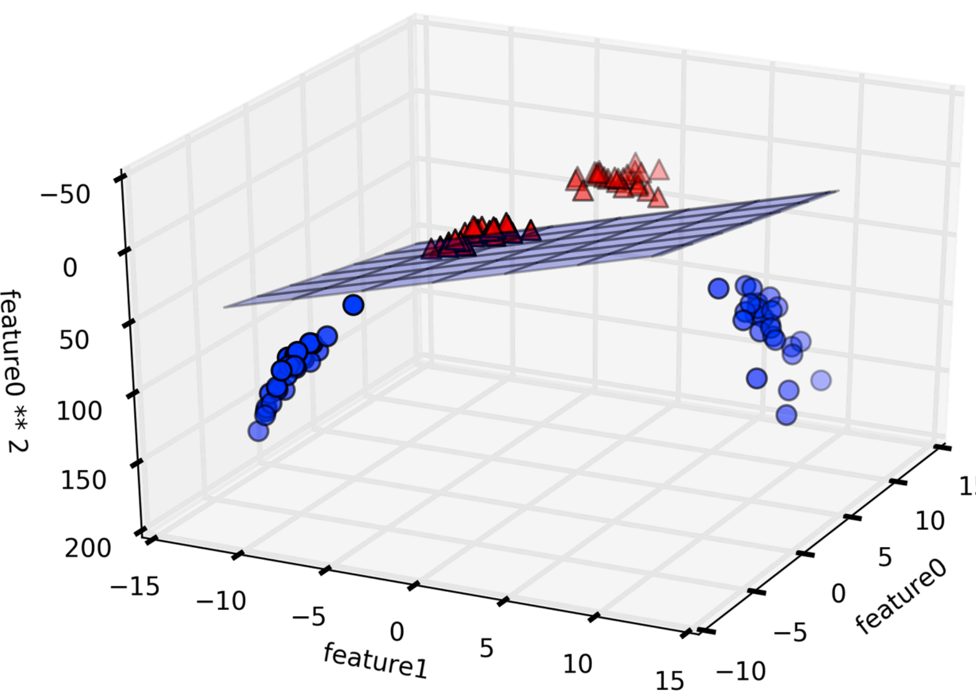


Figure 2-39. Decision boundary found by a linear SVM on the expanded three-dimensional dataset

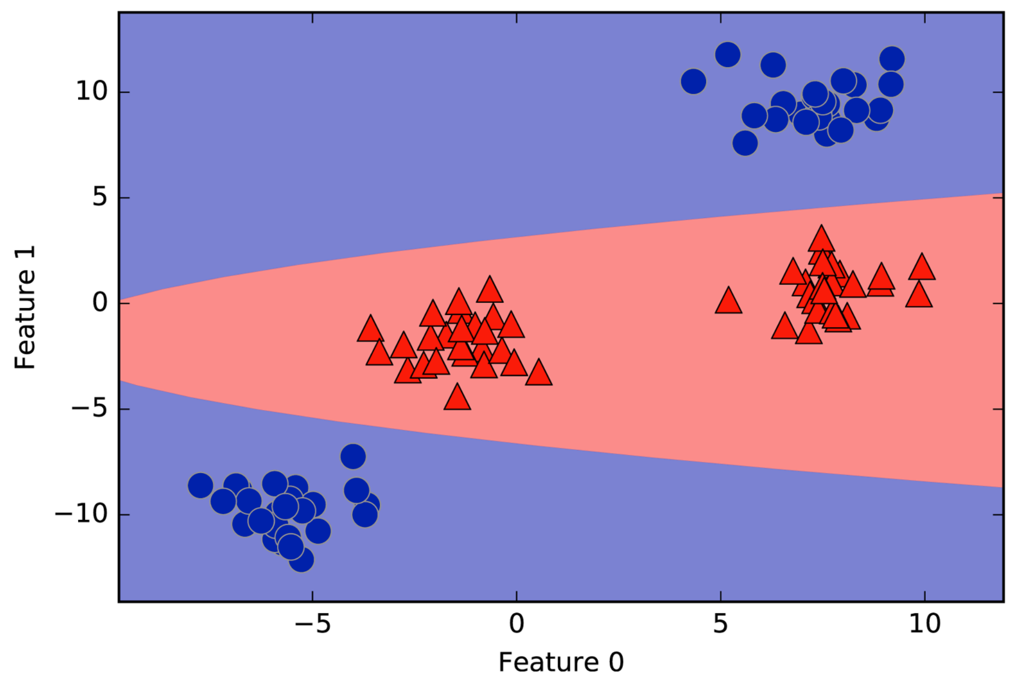


Figure 2-40. The decision boundary from Figure 2-39 as a function of the original two features

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.

Support Vector Machines

- Adding nonlinear features to the data can be useful
- Kernel trick - computes the distances (scalar products) between the nonlinear features (rather than having to calculate the full representation)
  - Polynomial kernel - all possible polynomials up to a specified degree
  - Radial basis function (RBF) - Gaussian weighting of all possible polynomials

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.

Support Vector Machines

- **Distance for the RBF is measured with the equation**

$$k_{\mathrm{rbf}}(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2)$$

  - **x_1 and x_2 are the data points**
  - **gamma sets the width of the Gaussian kernel**
    - **width of influence each data point**

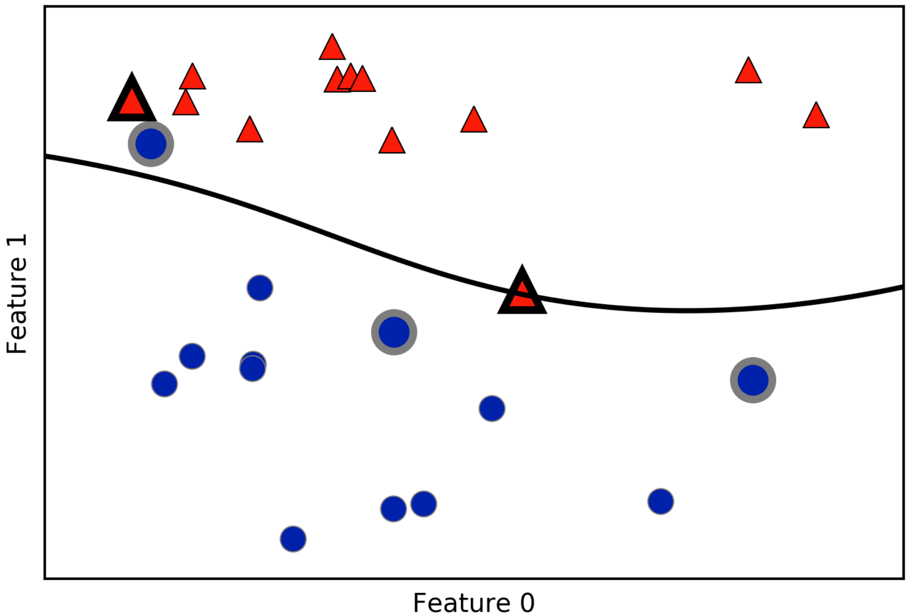


Figure 2-41. Decision boundary and support vectors found by an SVM with RBF kernel

- **As in linear models, the regularization is set by the C parameter**
  - **low values: algorithm adjusts to the majority of the data points**
  - **high value: algorithm attempts to correctly classify as many data points as possible**

Support Vector Machines

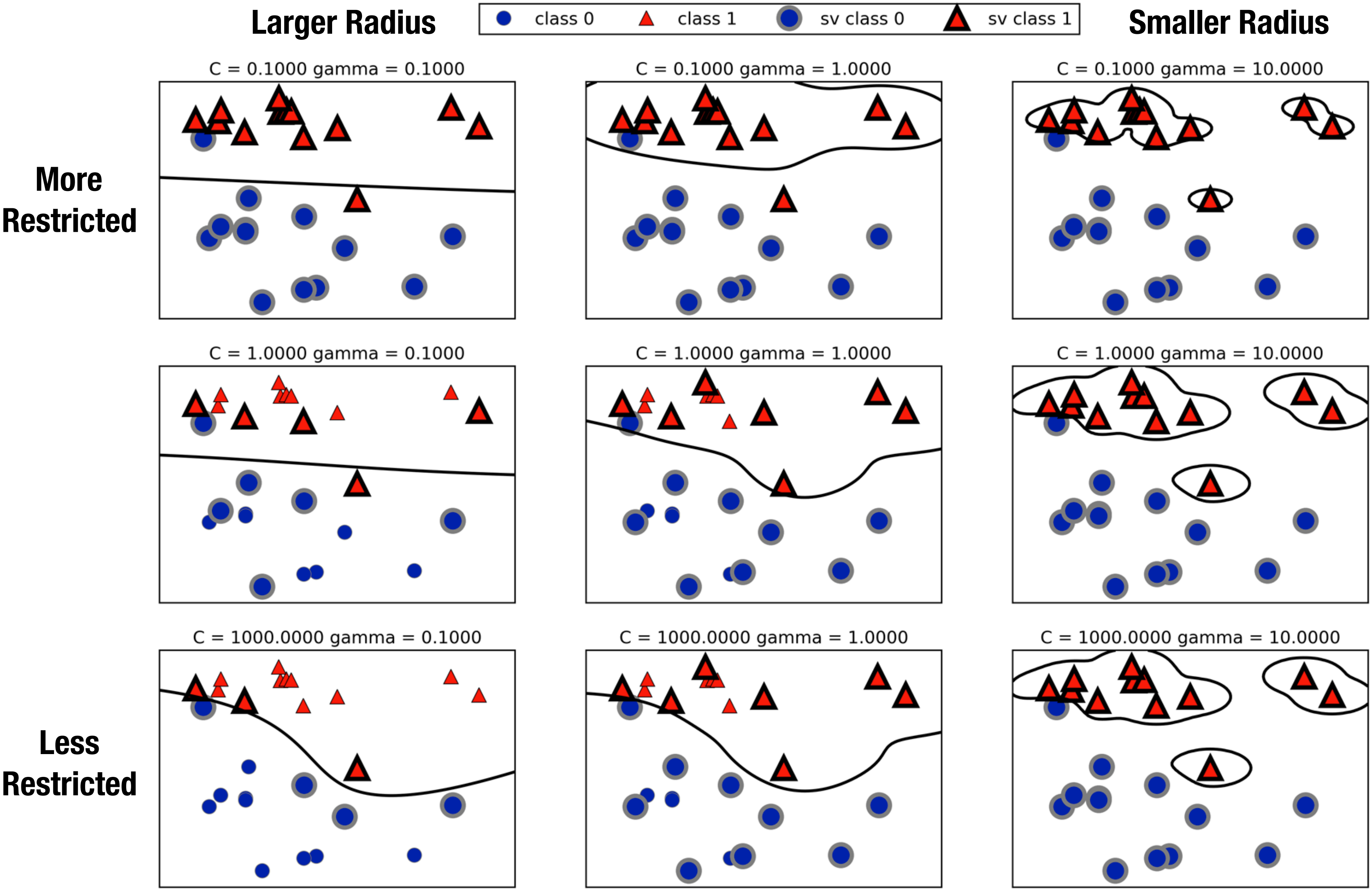


Figure 2-42. Decision boundaries and support vectors for different settings of the parameters C and gamma

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.

Support Vector Machines

- **Best Practices**
  - Features should be scaled to have 0 mean and unit variance
- **Parameters**
  - Choice of kernel (polynomial, radial basis function, etc.)
  - C - regularization parameter
  - gamma - inverse of width of Gaussian kernel (RBF)
- **Strengths**
  - Works well with only a few data points
  - Works on both low- and high-dimensional data
- **Weaknesses**
  - Doesn’t scale very well to a large number of samples
  - Data often needs to be pre-processed

Uncertainty Estimates from Classifiers

- **Decision Functions encode how strongly the model believes a given datapoint belongs to a given class**

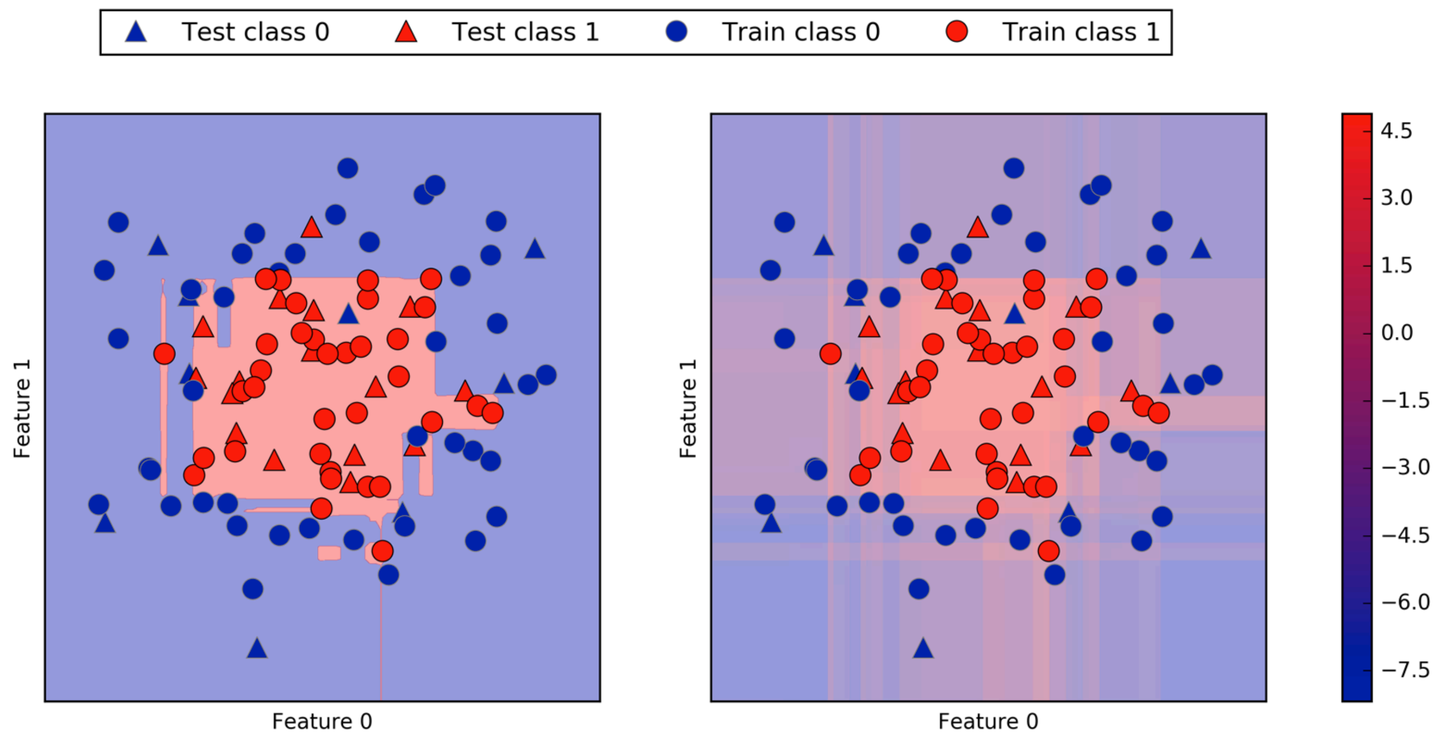


Figure 2-55. Decision boundary (left) and decision function (right) for a gradient boosting model on a two-dimensional toy dataset

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.

Uncertainty Estimates from Classifiers

- **Prediction Probabilities indicate the probability (between 0 and 1) that the a given data point belongs to a given class**

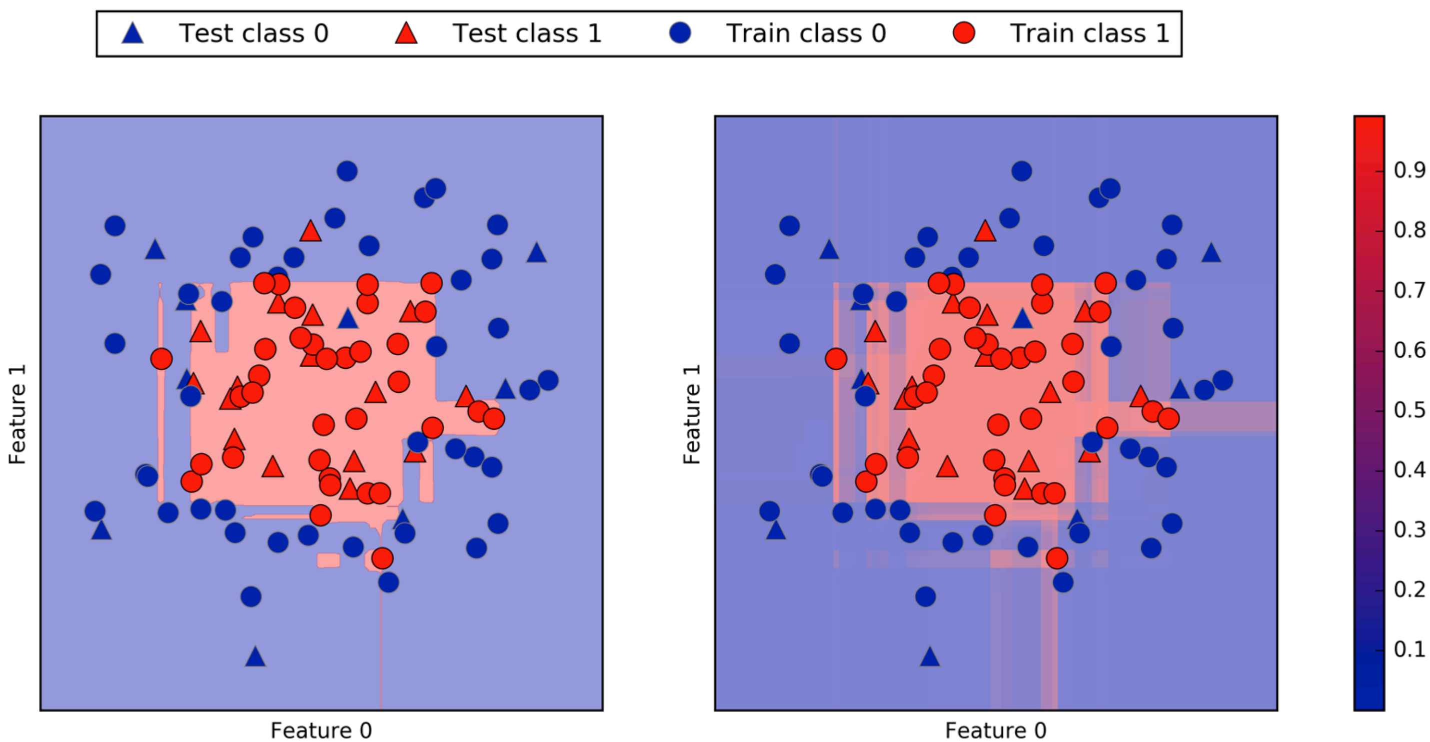


Figure 2-56. Decision boundary (left) and predicted probabilities for the gradient boosting model shown in Figure 2-55

Guido, Sarah and Andreas C. Muller. (2016). Introduction to Machine Learning with Python, O'Reilly Media, Inc.