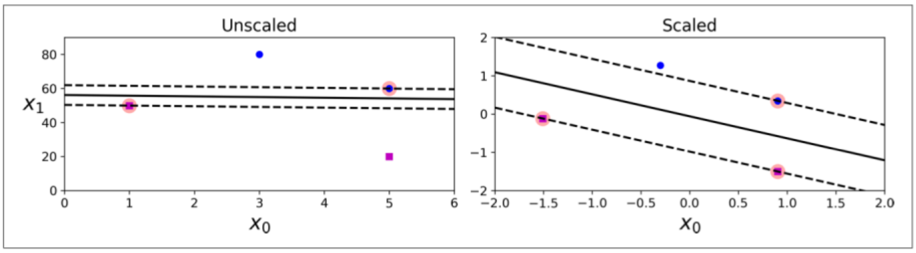
Support Vector Machines

Target: classification of complex but small- or medium-sized datasets

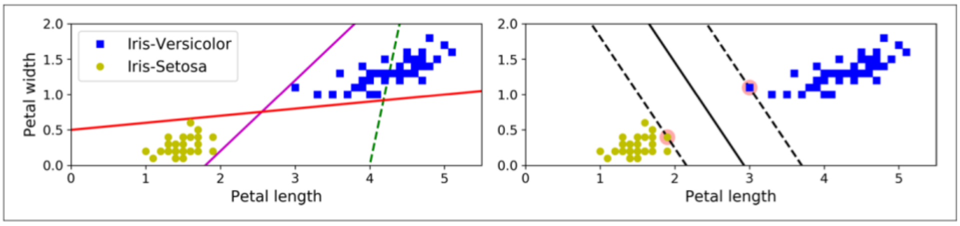
Goal: Find **decision boundaries** (hyperplane) for dataset

Notice: SVMs are sensitive to the feature scales (Scikit-Learn’s StandardScaler)

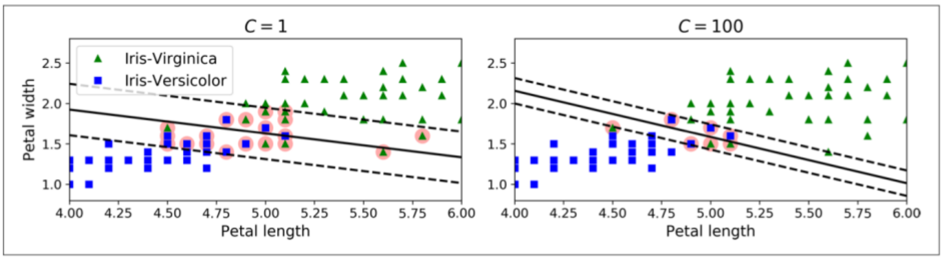


Types:

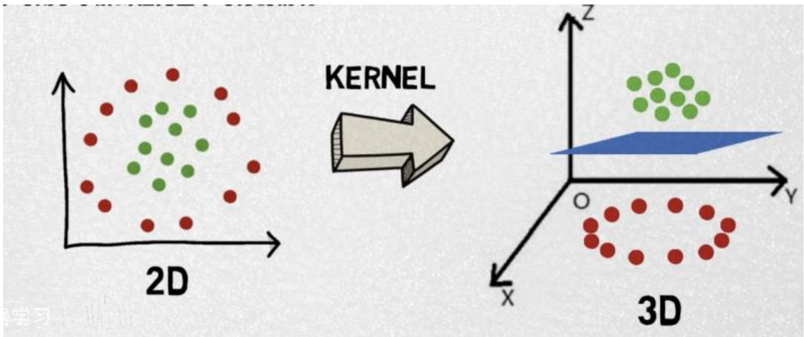
- **Linear SVM Classification**
    - **Support vectors**: decision boundary is fully determined by the instance located on the edge of the street (circled dot)
    - **Large margin classification**: find the widest possible street (dashed lines) between the classes, all instances should be off the street
        - **Two issues**:
            - data should be linearly separable
            - quite sensitive to outliers



-
    - **Soft margin classification:** find a good balance between keeping the street as large as possible and limiting the margin violations(i.e., instances that end up in the middle of street or even on the wrong side)
        - ***C*** hyperparameter: control the balance, a smaller C value leads to a wider street but more margin violations

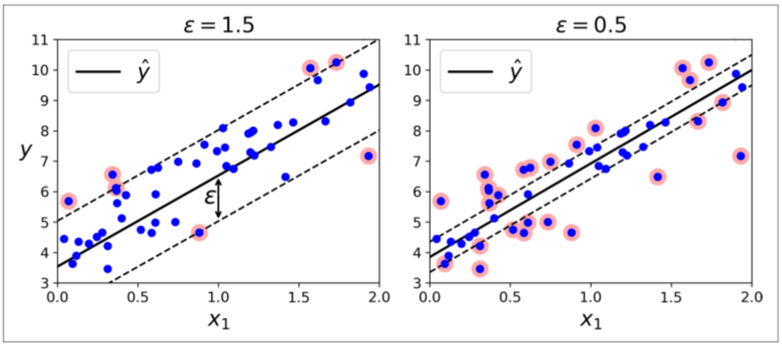


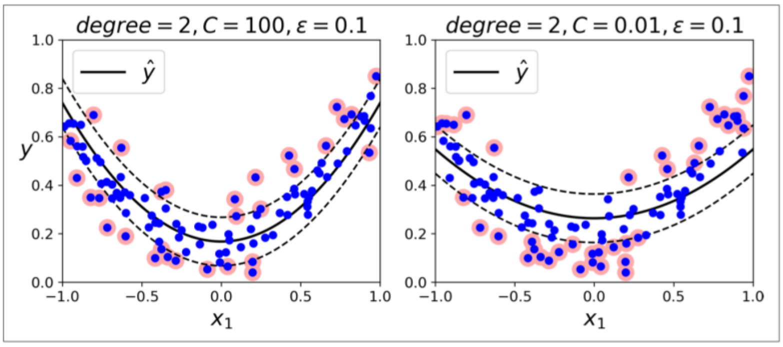
- **Nonlinear SVM Classification**
    - **Kernel trick:** project the instances in low dimensional space to those in high dimensional space



- ○ **Polynomial Kernel:** as if you added many polynomial features, but without acutally having added them
    - ▪ **Hyperparameter**
        - degree: polynomial degree
        - coef0: contorls how much the model is influecned by high-degree polynomials versus low-degree polynomials
- ○ **Gaussian RBF Kernel**: add features computed using a similarity function that measure how much each instance resembles a particular landmark
    - ▪ **Hyperparameter**
        - **Gamma (regularization term):** a larger value makes each instances's range of influence smaller -> the decision boundary ends up more irregular, wiggling around individual instances

- **SVM Regression**
    - ○ **essense**: fit as many instances as possbile on the street while limiting margin violation (i.e., instances off the street)
    - ○ **hyperparameter**:
        - ▪ ε: control the width of the street
        - ▪ C: regularization term


$\varepsilon = 1.5$
$\varepsilon = 0.5$
$\hat{y}$
y
ε
x_1


$degree = 2, C = 100, \varepsilon = 0.1$
$degree = 2, C = 0.01, \varepsilon = 0.1$
$\hat{y}$
y
x_1