GENERAL ASSEMBLY

SUPPORT VECTOR MACHINES

Joseph Nelson, Data Science Immersive

AGENDA

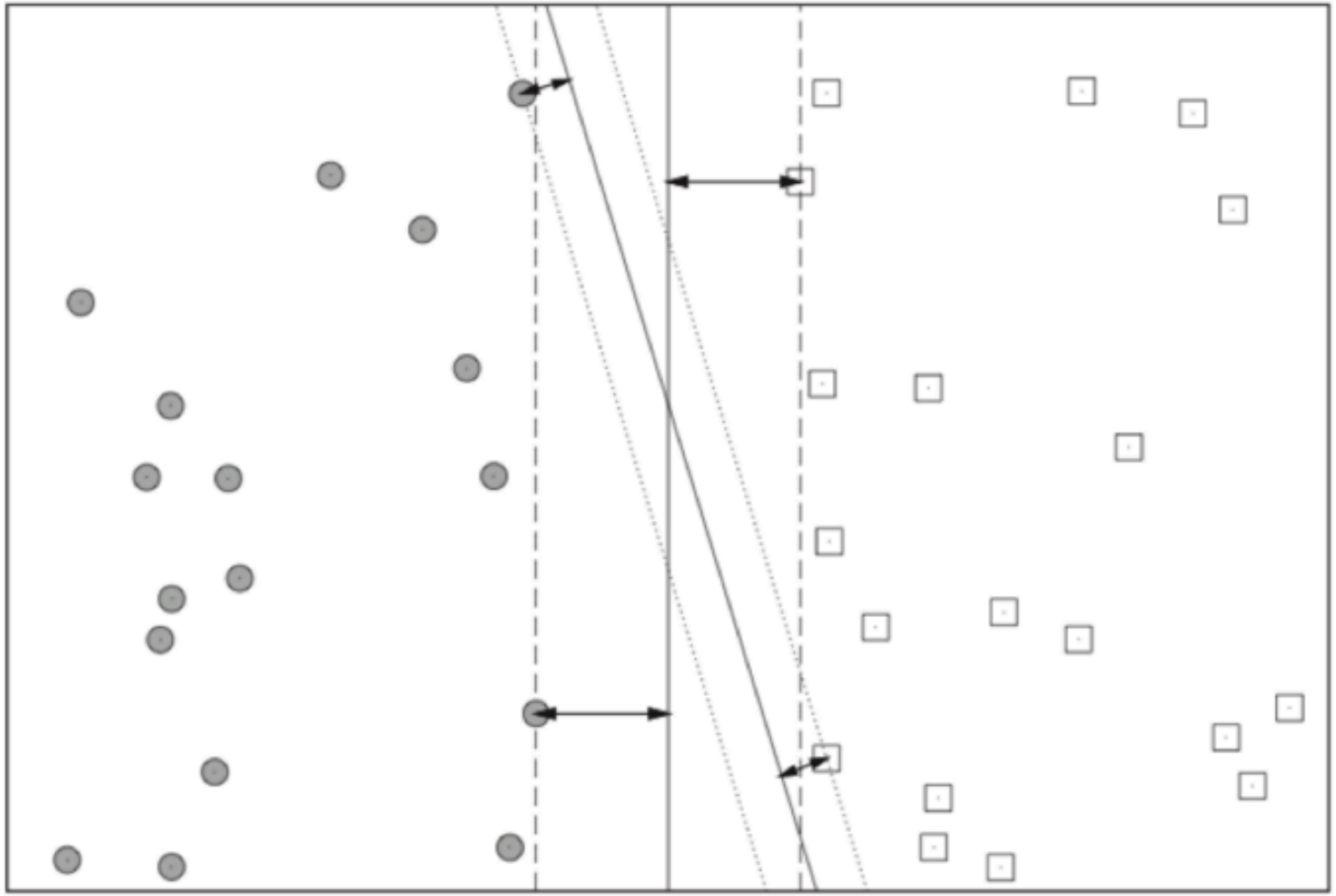
- What is a Support Vector Machine?
- How SVM Works
- Nonlinear SVM (Different Kernels)
- Coding Implementation
- Sklearn documentation

WHAT IS A SUPPORT VECTOR MACHINE?

- A Support Vector Machine is a binary linear classifier whose decision boundary is explicitly constructed to minimize generalization error
- Binary classifier: solves a two-class problem
- Linear classifier: creates a linear decision boundary

WHAT IS A SUPPORT VECTOR MACHINE?

- The decision boundary is derived using geometric reasoning (as opposed to the algebraic reasoning we've used to derive other classifiers). The generalization error is equated with the geometric concept of margin, which is the region along the decision boundary that is free of data points.



WHAT IS A SUPPORT VECTOR MACHINE?

- The goal of an SVM is to create the linear decision boundary with the largest margin. This is commonly called the maximum margin hyperplane (MMH).
- Nonlinear applications of SVM rely on an implicit (nonlinear) mapping that sends vectors from the original feature space K into a higher-dimensional feature space K’. Nonlinear classification in K is then obtained by creating a linear decision boundary in K’. In practice, this involves no computations in the higher dimensional space, thanks to what is called the kernel trick.

WHAT IS A SUPPORT VECTOR MACHINE?

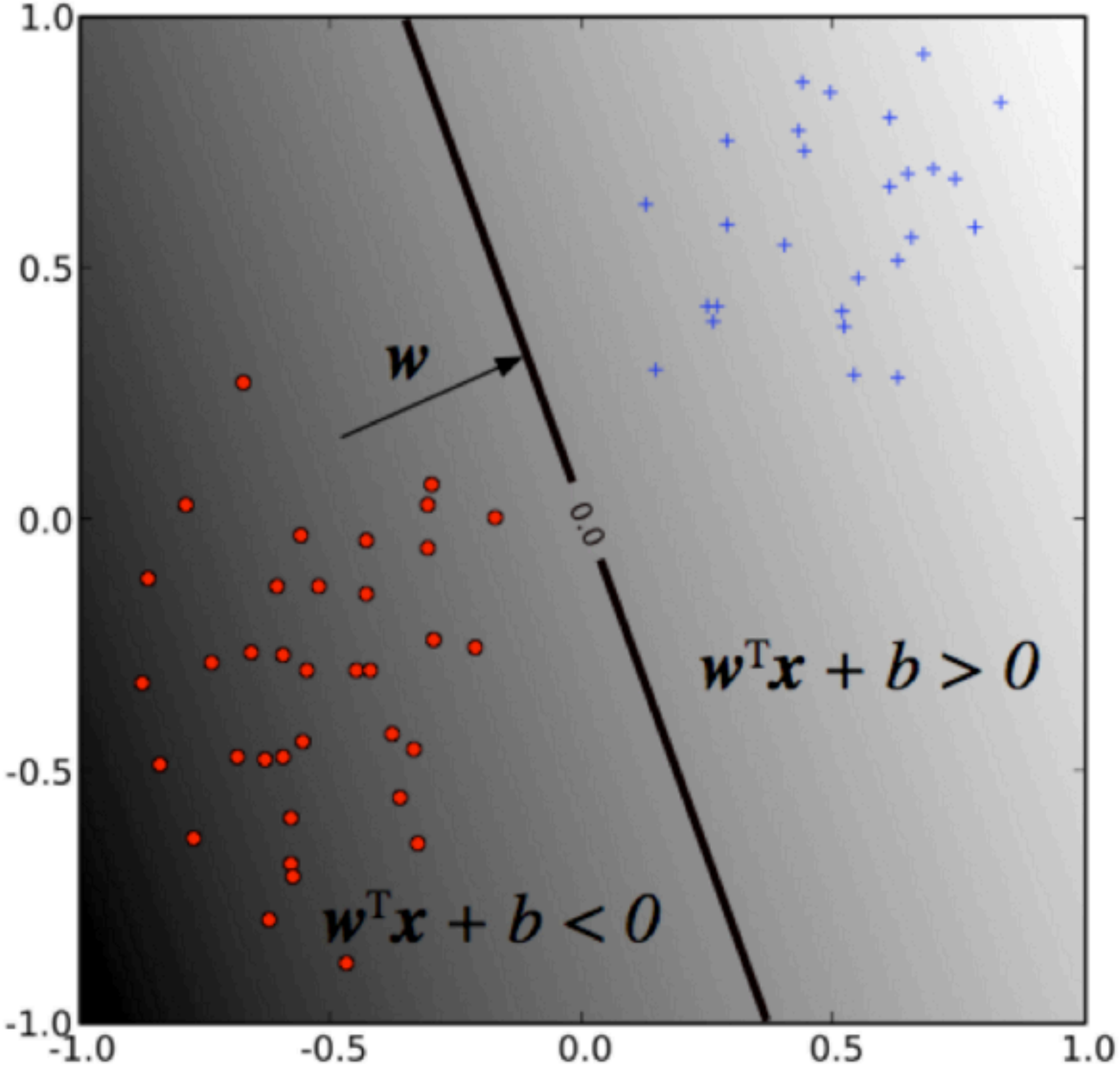
- The goal of an SVM is to create the linear decision boundary with the largest margin. This is commonly called the maximum margin hyperplane (MMH).
- Nonlinear applications of SVM rely on an implicit (nonlinear) mapping that sends vectors from the original feature space K into a higher-dimensional feature space K'. Nonlinear classification in K is then obtained by creating a linear decision boundary in K'. In practice, this involves no computations in the higher dimensional space, thanks to what is called the kernel trick.

DECISION BOUNDARY

- The decision boundary (MMH) is derived by the discriminant function:

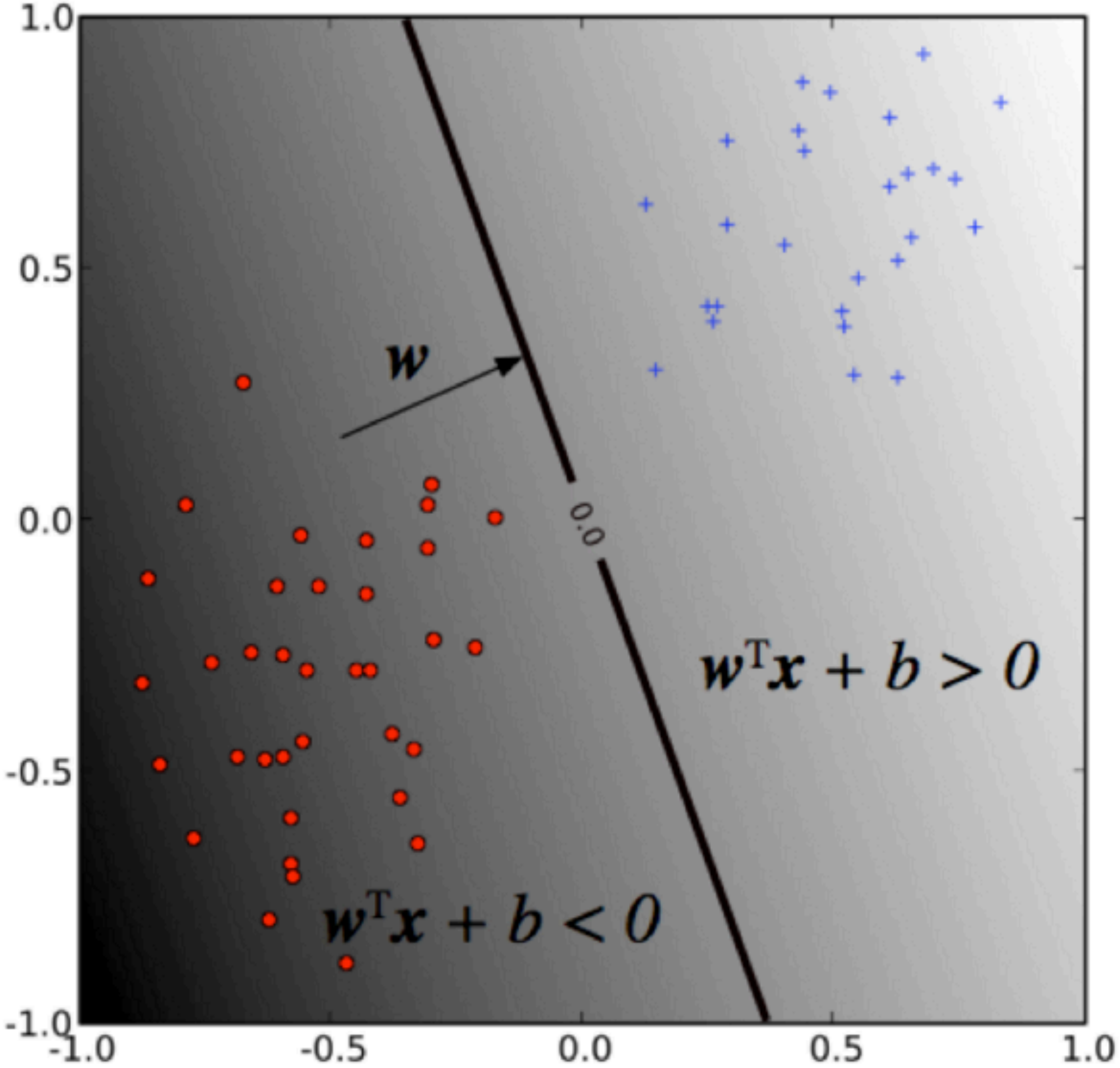
f(x) = w^T x + b

- where w is the weight vector and b is the bias. The sign of f(x) determines the (binary) class label of a record x.



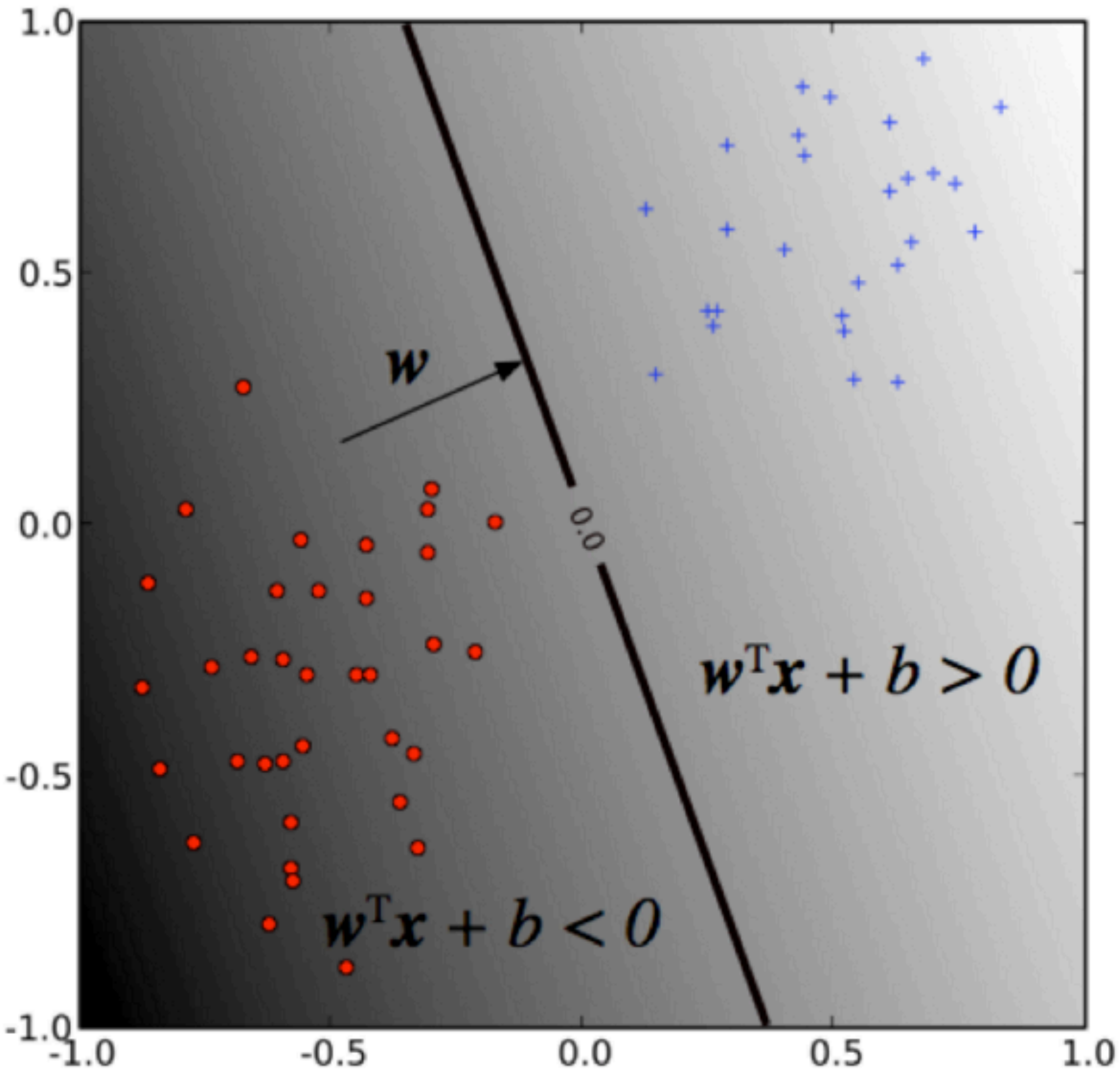
DECISION BOUNDARY

- Remember: SVM solves for the decision boundary that minimizes generalization error, or equivalently, that has the maximum margin. These are equivalent since using the MMH as the decision boundary minimizes the probability that a small perturbation in the position of a point produces a classification



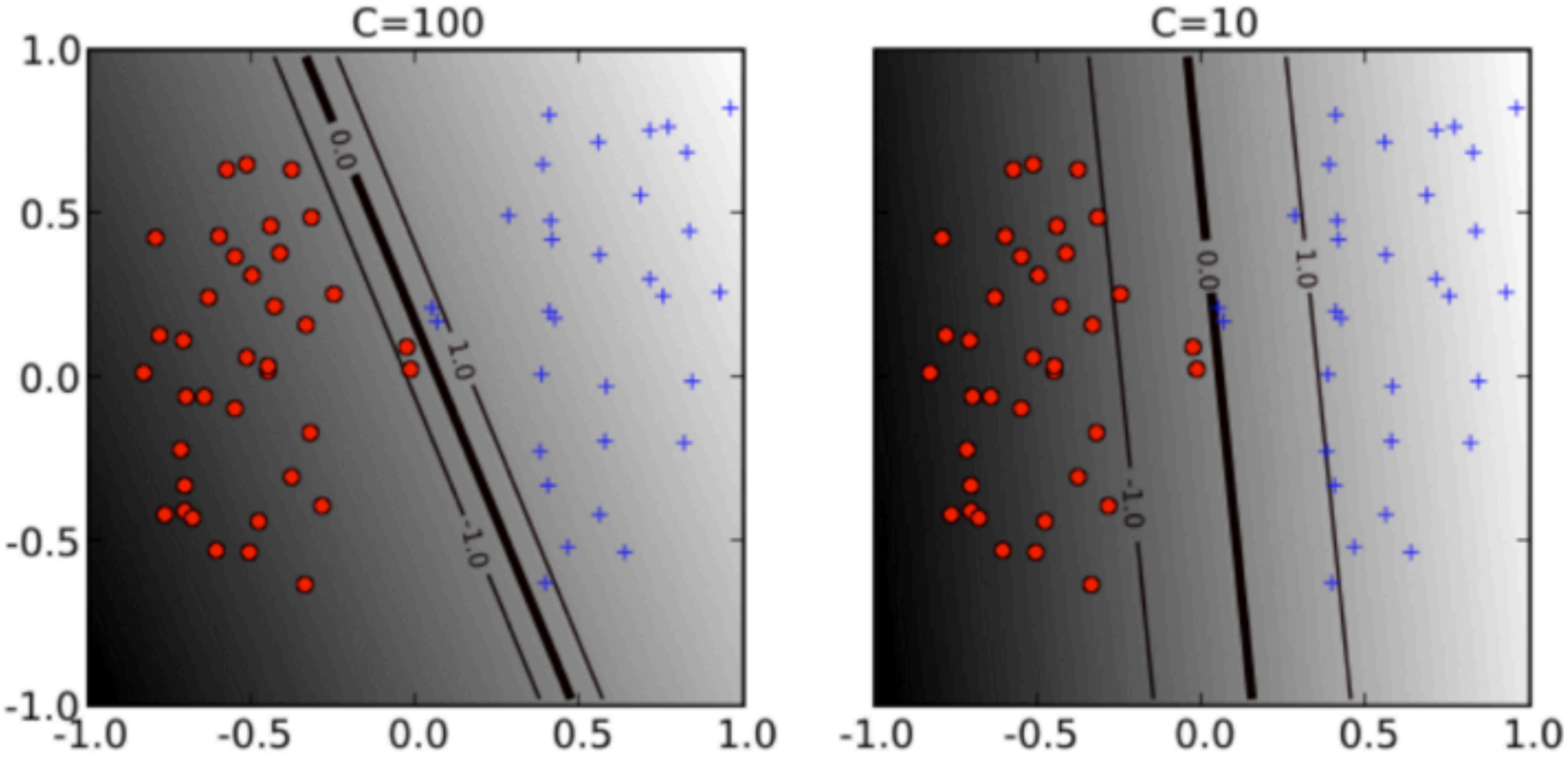
DECISION BOUNDARY

- Selecting the MMH is a straightforward exercise in analytic geometry
- The margin depends only on a subset of the training data; namely, those points that are nearest to the decision boundary. These points are called the support vectors. The other points (far from the decision boundary) don't affect the construction of the MMH at all.



SOFT MARGIN, SLACK VARIABLES

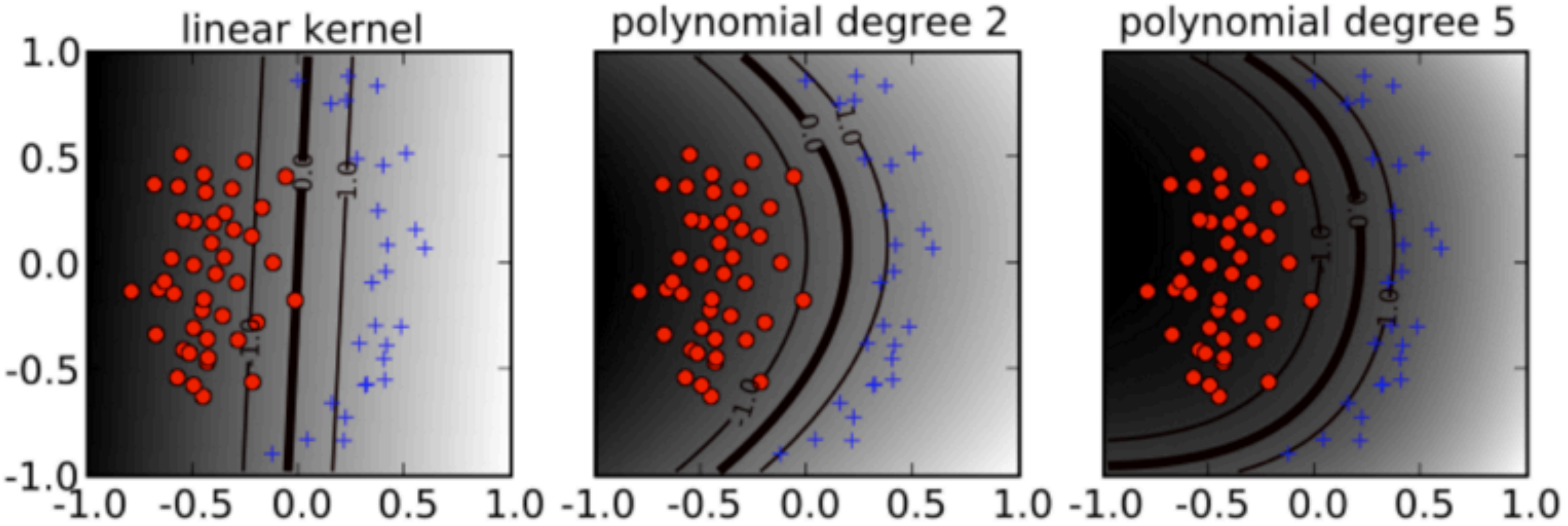
- Class overlap is achieved by relaxing the minimization problem or softening the margin.
- The hyper-parameter C (soft-margin constant) controls the overall complexity by specifying penalty for training error. This is yet another example of regularization.



NONLINEAR SVM

- At its core, the optimization problem only has x as an inner product
- $f(x) = w^T x + b$
- We can replace this inner product with a more complex function: this is called the kernel trick
- Popular kernels:
- Linear $k(x, x') = x^T x'$
- Polynomial $k(x, x') = (x^T x' + 1)^d$
- Gaussian kernel (radial basis function) $k(x, x') = \exp{(-\gamma||x - x'||^2)}$

NONLINEAR SVM



NONLINEAR SVM