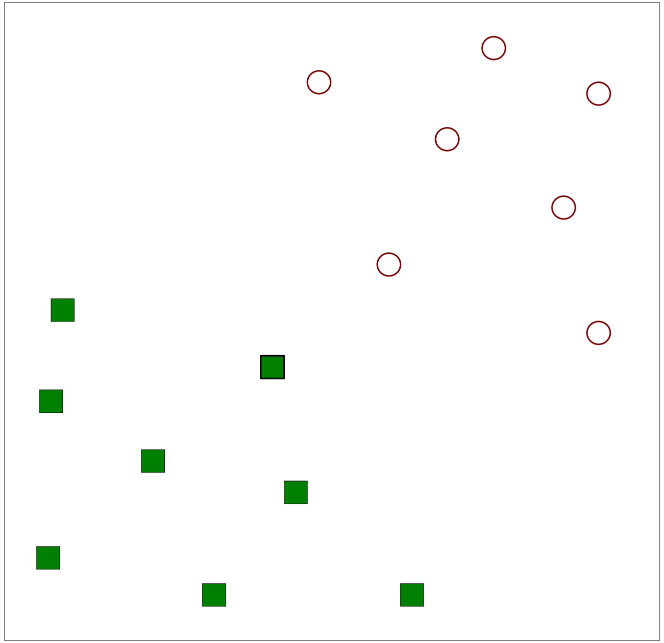
Linear Support Vector Machines (SVM)

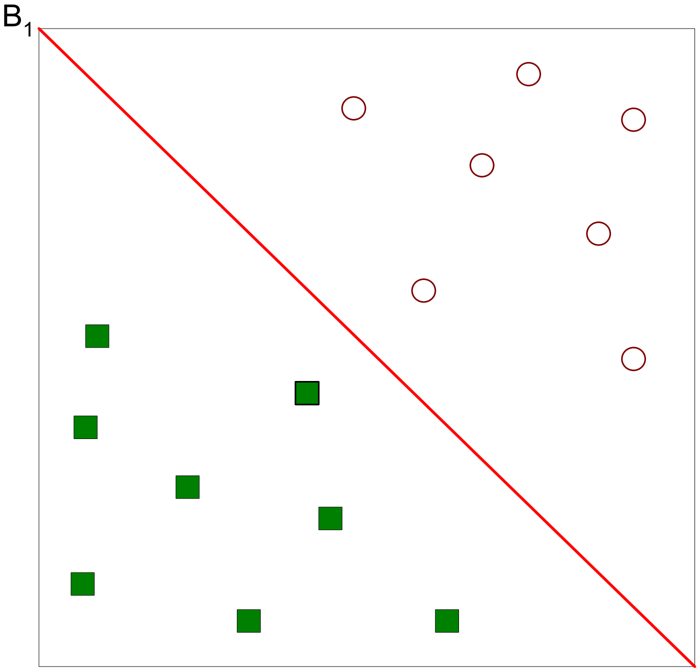
Part II

Support Vector Machines: Part I



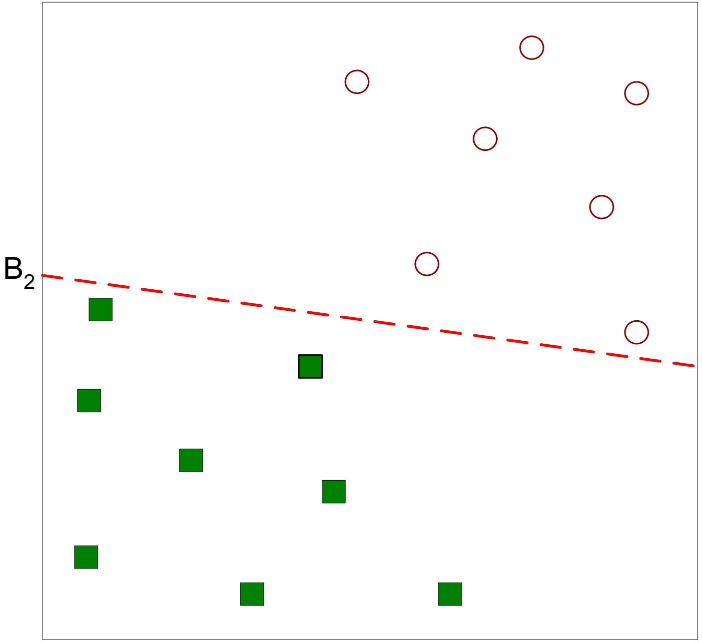
Find a linear hyperplane (decision boundary) that will separate the data

Support Vector Machines: Part II



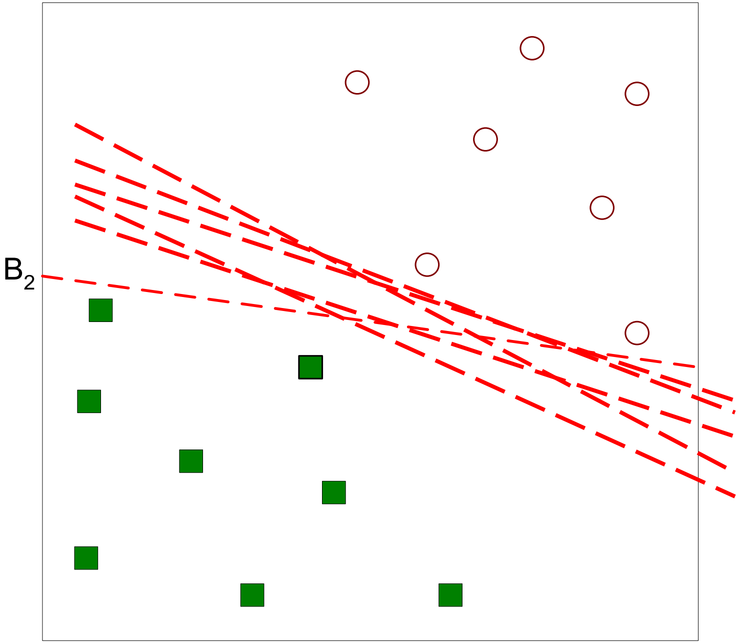
One possible solution

Support Vector Machines: Part III



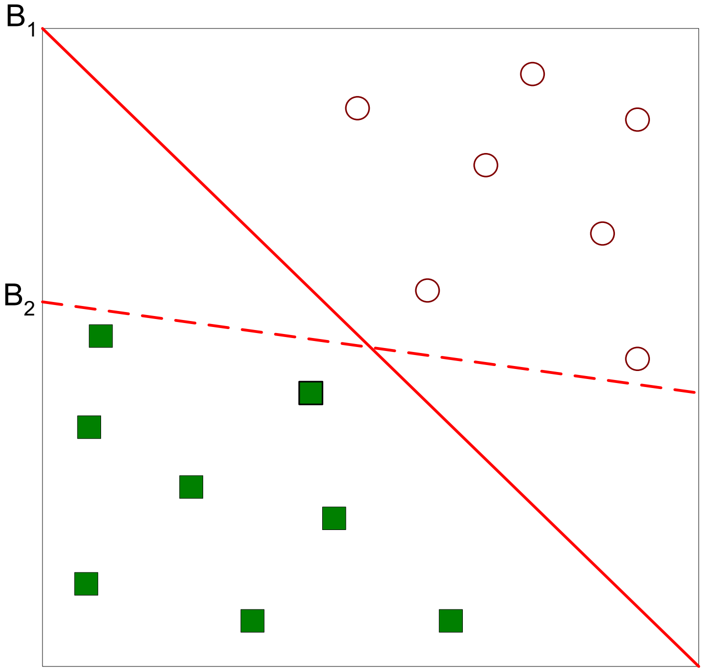
Another possible solution

Support Vector Machines: Part IV



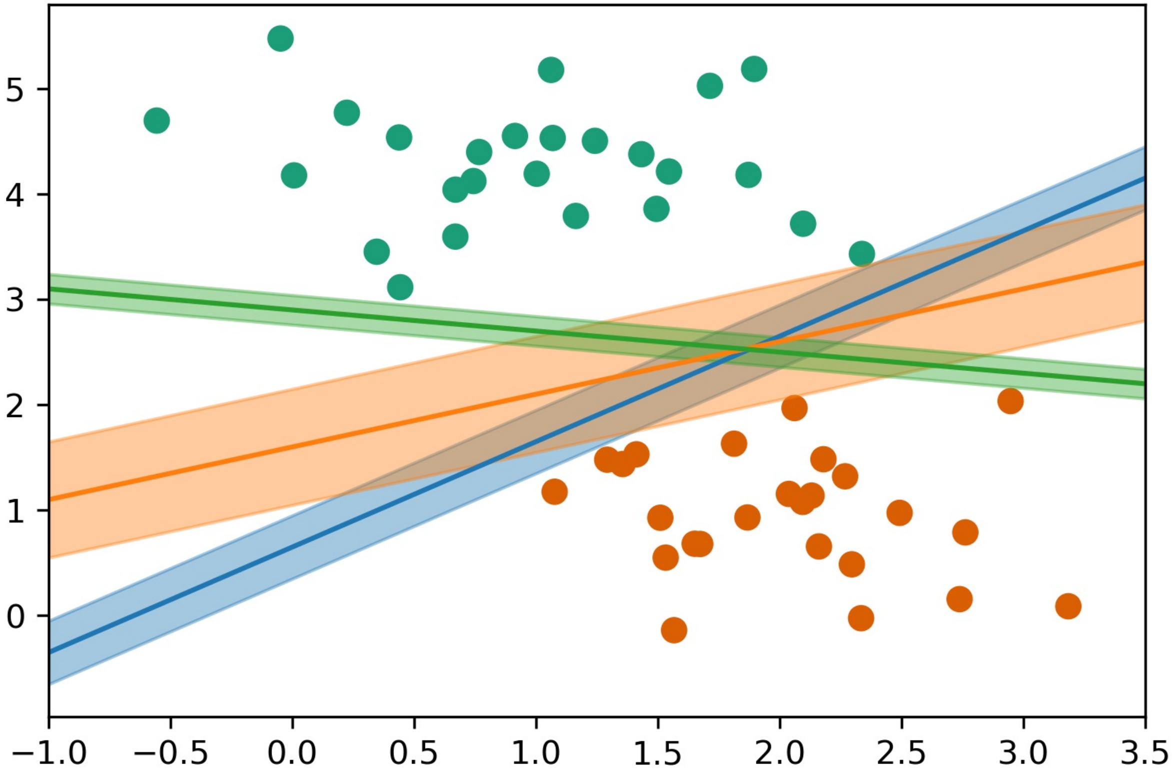
Other possible solutions

Support Vector Machines: Part V



- Which one is better? B1 or B2?
- How do you define better?

Support Vector Machines: Part VI



Decision Boundary

Constraints on the decision boundary:

- In logistic regression, we typically learn an $\ell 1$ or $\ell 2$ regularized model.
- So, when the data is linearly separable, we choose a model with the *smallest coefficients* that still separate the classes.
- The purpose of regularization is to prevent overfitting.

Decision Boundary, Continued

- We can consider alternative constraints that prevent overfitting.
- For example, we may prefer a decision boundary that does not ‘favor’ any class (especially when the classes are roughly equally populous).
- Geometrically, this means choosing a boundary that maximizes the distance or ***margin*** between the boundary and both classes.

Maximizing Margins

- Notice that maximizing the distance of ***all points*** to the decision boundary is exactly the same as maximizing the distance to the ***closest points***.
- The points closest to the decision boundary are called ***support vectors***.
- For any plane, we can always scale the equation:

$$w^{\mathsf{T}}x + b = 0$$

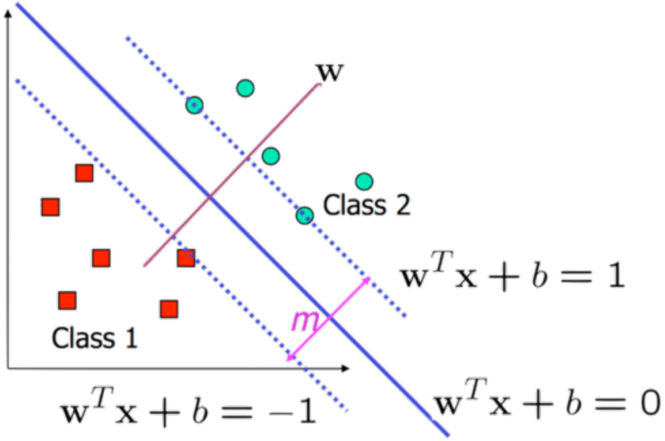
so that the **support vectors** lie on the planes:

$$w^{\mathsf{T}}x + b = \pm 1$$

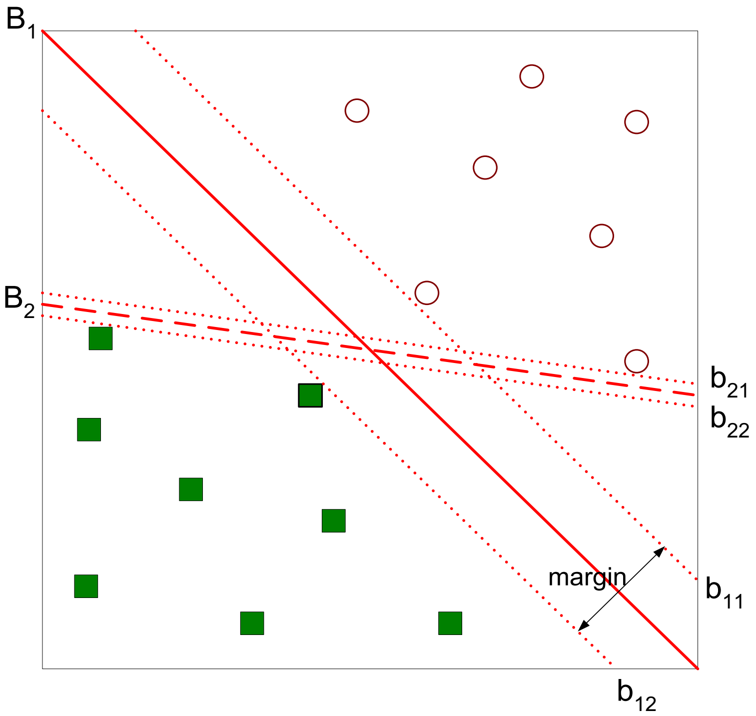
depending on their classes.

Maximizing Margins Illustration

- For points on planes $w^T x + b = \pm 1$, their distance to the decision boundary is $\pm 1/\|w\|$.
- So we can define the ***margin*** of a decision boundary as the distance to its support vectors, $m = 2/\|w\|$.

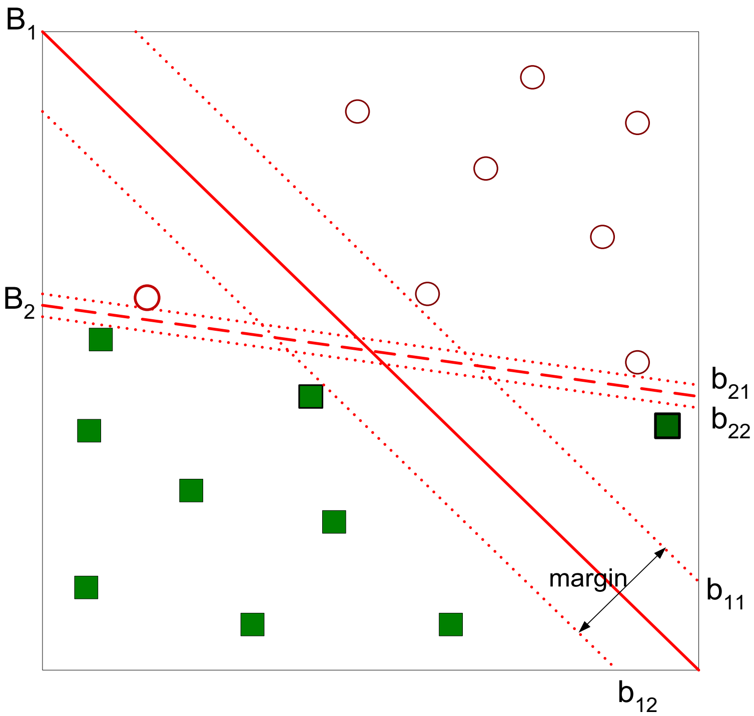


Support Vector Machines Large Margin Classifier (Hard Margin)



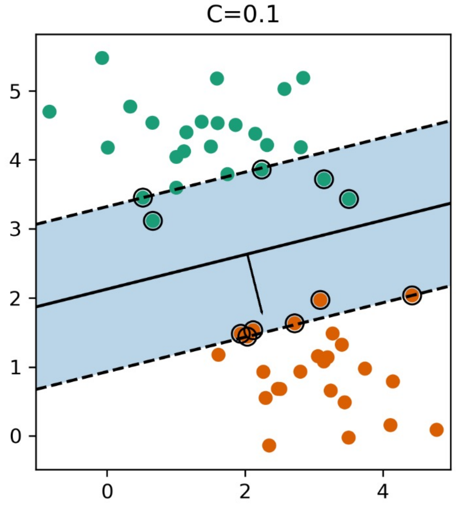
Find the hyperplane that **maximizes** the margin => B1 is better than B2

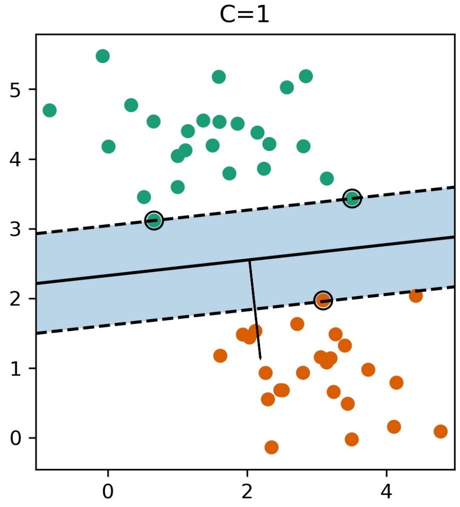
Support Vector Machines Large Margin Classifier (Soft Margin)



Find the hyperplane that **optimizes** both factors

Regularization for SVM: The C parameter (Soft Margin)





Linear Models: Pros and Cons

Pros

- Simple and easy to train
- Fast prediction
- Scales well with very large datasets
- Works well with sparse data
- Reasons for prediction are relatively easy to interpret

Cons

- For lower-dimensional data, other models may have superior generalization performance.
- For classification, data may not be linearly separable (more on this in SVMs with non-linear kernels).