ECE M146 Introduction to Machine Learning

Prof. Lara Dolecek

ECE Department, UCLA

Today’s Lecture

Recap:

- Support Vector Machines (SVM)

New topic:

- Kernels
- Soft SVM

Today’s Lecture

Recap:

- Support Vector Machines (SVM)

New topic:

- Kernels
- Soft SVM

Recap: SVM

- Last time we introduced the SVM method for binary classification.
- This method finds the vector w so that the smallest distance from a training point to the decision boundary is maximized.
- Picture:

Recap: SVM

- We then formulated the minimization problem as a quadratic program with linear constraints.
- A way to solve such a problem is by converting a constrained optimization problem into an unconstrained optimization problem.
- This is done with a Lagrangian, which incorporates constraints with Lagrange multipliers.

Recap: SVM

- Note that we are here solving a minimization problem.
- Utility of the Lagrangian extends to maximization problems as well as minimization/maximization problems with inequality and/or equality constraints.
    - The difference is in the sign.

Recap: SVM

- Even though the we could have solved the unconstrained problem as stated (primal), we intentionally converted it into a different problem (dual).
- Math:
- The advantage of the dual formulation is that all the terms are scaled versions of $x_n^T x$.

Today's Lecture

Recap:

- Support Vector Machines (SVM)

New topic:

- Kernels
- Soft SVM

Achieving linear separability in higher dimensions

- Consider the following example:

- If we use the original 2-d vector representation, the data is clearly not linearly separable.
- Now, consider, the following map $\varphi(x)$ of a data point x:

Achieving linear separability in higher dimensions

- By adding new features, we made the classes linearly separable.
- But does this only mean that we need to do more computations ?
- Let’s see.

Feature expansion example

- Let’s consider the following example. Suppose x and z are given as:

- Now suppose that their maps are as follows:

- We went from having 2 to having 7 dimensions!

Example, continued

- Compute the inner product $\varphi(x)^{\mathsf{T}}\varphi(z)$:

- Compare to $\left(1 + x^T z\right)^2$:

- Same answer, but the former is much faster. Kernel trick!

Kernels

- There are many valid kernels, beyond example we just saw.
- Polynomial kernel:
    - Previous example was for d=2.
- Gaussian kernel:

When does the map $\varphi(x)$ produce a valid kernel ?

- Suppose we have N data points.
- Let K_{ij} be the following:

- Note that with this representation we do not need to specify the map explicitly!
- The overall matrix K is NxN and symmetric.

Properties of this matrix

- Consider the following:

- Therefore, the matrix is positive semi-definite (PSD).

Example

- Let's consider the Gaussian kernel again, for 2-d vectors:

- Show PSD property:

Example, continued

Another example

- Now consider the following example

More on kernels

Kernels in SVM

Today's Lecture

Recap:

- Support Vector Machines (SVM)

New topic:

- Kernels
- Soft SVM

Soft margin SVM

- Recall SVM for linearly separable data:

- But if the data is (almost) not linearly separable, we are forcing a decision boundary that is not good – poor generalization error.

Instead, let’s allow for slack

- New linear constraints:

- Case 1:

- Case 2:

- Case 3:

Instead, let’s allow for slack

- New linear constraints:

- Case 4:

- Case 5:

Soft margin SVM

- Formulation:

- Role of the hyperparameter C:

Support vectors in soft SVM

- Now, the support vectors will be all the points within the margin boundary and on it.

- There are typically more support vectors in soft SVM than in the original (hard) SVM.

Reformulation of the Lagrangian

- Before we had:

- Now we have:

  - Lagrange multipliers:

KKT conditions

- Old conditions are still here:

- New conditions are:

Partial derivatives

- These are same as before:

- These are new:

Expression for the Lagrangian

- We can eliminate w, b, ε_{n} from the expression (similar to the case we had before) to obtain:

- Compact representation:

Support vectors

- If $a_n = 0$

- If $a_n \neq 0$

Support vectors

Relation to logistic regression

Relation to logistic regression

Extensions