Week 5: Support Vector Machines and Expectation-Maximization

March 1, 2019

1 Support Vector Machines

1.1 Support Vector Loss

The basic optimization objective for Support Vector Machines is

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\|w\| + C\sum_i p_i \\ \text{such that} \quad & y_i(w^T x_i + b) \geqslant 1 - p_i \\ \text{and} \quad & p_i \geqslant 0 \end{aligned}$$

Exercise 1

What does i index? How many terms does the sum have, and how many constraints are there?

What is the value of y_i in this expression? What is its function?

Exercise 2

There are two common ways to rewrite this expression before implementing it. What are they (in general terms) and what are their benefits?

1.2 Lagrange multipliers

Lagrange multipliers are a useful trick to know. We'll practice them briefly on a small problem, so that you understand the principle. We will only use an *equality* constraint.

We have the following optimization problem:

$$\begin{aligned} \text{minize } f(a, b) &= a^2 + 2b^2 \\ \text{such that } a^2 &= -b^2 + 1 \end{aligned}$$

Exercise 3

The first step is to rewrite the constraint so that the right side is equal to zero. Do so.

What does the constraint say about the allowed inputs (what shape do the allowed inputs make in the (a, b)-plane)?

We now define a function $L(a, b, \lambda) = f(a, b) + \lambda G$, where G is the left hand side of the constraint equal to zero (how much any given a and b violate the constraint).[1]

Exercise 4

Write out $L(a, b, \lambda)$ for our problem.

We take the derivative of L with respect to each of its *three* parameters, and set these equal to zero.

Exercise 5

Fill in the blanks

$$a(\ldots) = 0 \tag{1}$$

$$b(\ldots) = 0 \tag{2}$$

$$a^2 + b^2 = 1 \tag{3}$$

Note that the last line recovers the original constraint. We now have three equations with three unknowns, so we can solve for a and b. From the shape of the function (it's symmetric in both the a and b axes), we should expect at least two solutions.

We can get these from the above equations by noting that if a and b are both nonzero, we can derive a contradiction. Thus either a or b must be zero.

[1]For plain Lagrange multipliers, where the constraints are all equalities, we can either add or subtract the term containing the constraint. For inequality constrains, it depends on whether we are maximizing or minimizing.

Exercise 6

Give the solutions for both cases (remember that $x^2 = 1$ has *two* solutions).

Happily, Wolfram Alpha agrees with us (and provides some informative plots).

1.3 The kernel trick

The feature space of k is a projection of point $\mathbf{a}$ to point $\mathbf{a}'$ such that

$$k(\mathbf{a}, \mathbf{b}) = \mathbf{a}'^{\mathsf{T}}\mathbf{b}' \,.$$

We have a dataset with two features Let $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$. We define the *kernel*

$$k_1(\mathbf{a}, \mathbf{b}) = (\mathbf{a}^{\mathsf{T}}\mathbf{b})^2 \,.$$

Exercise 7

Show that the feature space defined by$k_!$ is

$$\begin{pmatrix} {x_1}^2 \\ \sqrt{2}x_1x_2 \\ {x_2}^2 \end{pmatrix} \,.$$

Hint: start by writing out the definition as a scalar function. See if your can re-arrange this back into a dot procut of two other vectors.

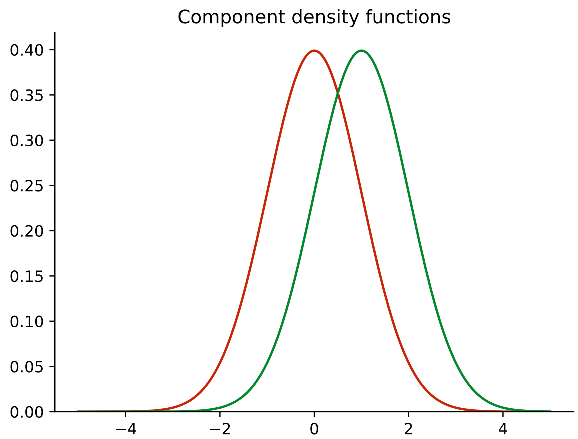
Exercise 8

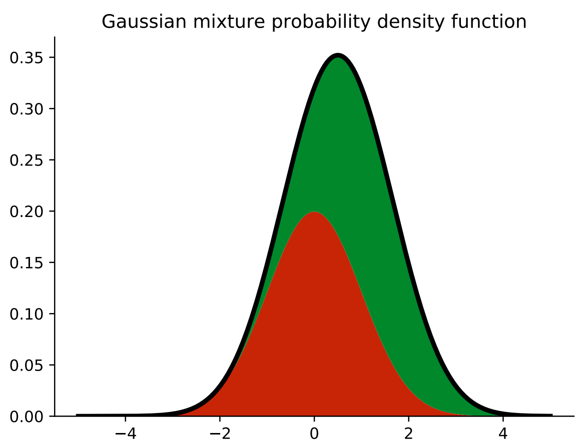
What is the feature space for the kernel

$$k_2(\mathbf{a}, \mathbf{b}) = \left(\mathbf{a}^{\mathsf{T}}\mathbf{b} + 1\right)^2 \quad ?$$

2 Expectation Maximization

Assume we have a Gaussian Mixture Model in one dimension with two components: $N(0, 1)$ and $N(1, 1)$. The weights w_1 and w_2 of the components are equal.





Exercise 9

Compute the probability density of the point 0, under the Gaussian Mixture.

Exercise 10

Under the EM algorithm, what responsibility is assigned to each component for the point 0?