what does scaling the normal vector of a plane (/hyperplane) mean?

Asked 2 years, 4 months ago Active 2 years, 4 months ago Viewed 879 times

2

2

I understand that, scaling (multiplying or dividing by a constant) the normal vector of a plane, does not affect the plane itself. But what happens when we do so? Are we zooming in or out of the space, like in a linear transformation?

Or in simpler words, what's the intuitive effect on scaling the normal vector of a plane(or a hyperplane)?

I came across the optimization problem whilst learning SVM (support vector machines) which goes like this: minimize $w^\top w$ s.t. $\forall_i,\ \ y_i(w^\top \cdot x_i + b) \geq 1$, where w is the normal vector of the supporting hyperplane. So we are basically looking to minimize the length of the normal vector of the hyperplane ($w^\top \cdot w$) subject to the constraint. So I'm just wondering what effect could come out of minimizing the length of the normal vector.

machine-learning svm linear-algebra

edited Mar 14 '17 at 23:56
Antoni Parellada
15.7k 8 65 149

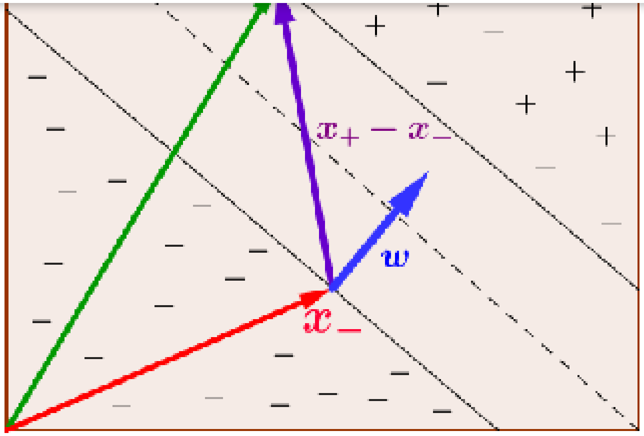
asked Mar 13 '17 at 21:22
Nithish Inpursuit Ofhappiness
275 3 8

2 Answers

2

We want to create as broad a separation (a wide "street") between positive and negative examples, or in other words, maximize the distance between the tips of the support vectors and the decision boundary (the median of the street):

SVM street



If the normal vector to the decision hyperplane, $\vec{w}$, is normalized, $\frac{\vec{w}}{\|\vec{w}\|}$, the street width will be equal to the dot product of a vector spanning the distance between any two points in the positive and negative boundary limits ("the gutter"), (x_+ and x_-), and $\frac{\vec{w}}{\|\vec{w}\|}$.

Imposing the constraint $y_i(w^\top \cdot x_i + b) \geq 1$ will have the positive effect of maximizing the width of the street; however b is not a predetermined value, and in fact, it is not part of the final Lagrangian expression. As for y_i, the objective is simply to keep the sign of the inequality (semi)-positive.

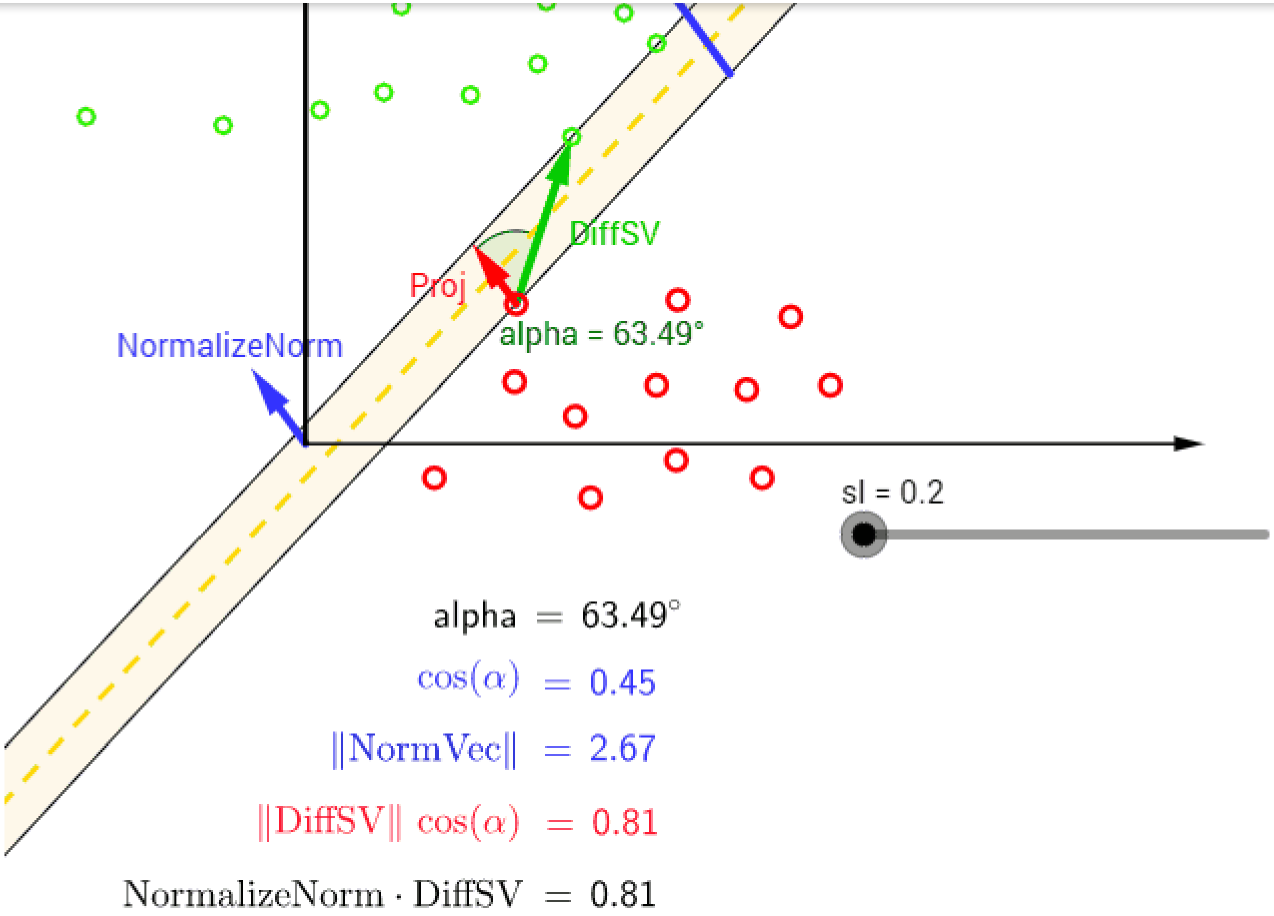
Given these premises, we can quickly arrive at the conclusion that maximizing the width of the "street" is equivalent to minimizing the norm of $\vec{w}$.

Above x_+ and x_- *are* in the gutter (on hyperplanes maximizing the separation). Therefore, for the positive example: $(\mathbf{x_i} \cdot w + b) - 1 = 0$, or $\mathbf{x}_+ \cdot w = 1 - b$; and for the negative example: $\mathbf{x}_- \cdot w = -1 - b$. So, reformulating the width of the street:

$$\text{width} = (x_+ - x_-) \cdot \frac{w}{\|w\|} = \frac{2}{\|w\|} \qquad \text{(the width of the street)}$$

We just have to maximize the width of the separation, which amounts to maximizing $\frac{2}{\|w\|}$, or minimizing $\|w\|$.

This can easily be verified analytically, or with a geometry game:



The positive examples are green dots, and the negative examples, red. Notice that by changing the slope of the parallel lines (the street gutters), we make the separation broader and broader until reaching a maximum value, and along the process, the norm or length of the normal vector (chosen randomly) decreases from $\|\text{NormVec}\| = 2.67$ to $\|\text{NormVec}\| = 2$. The vector spanning the difference between a positive and a negative example in the gutter is DiffSV (difference support vector), and its projection orthogonal to the decision boundary (Proj) (projection) increases in magnitude as the norm of the normal vector to the decision hyperplane decreases.

This makes sense, because at the same time, the angle α between DiffSV (difference support vector) and Proj (projection) decreases, and the

Here is a Geogebra toy example.

edited Mar 15 '17 at 14:05

answered Mar 14 '17 at 2:22

Antoni Parellada
15.7k 8 65 149

I think you are pointing me in the right direction, but i don't completely follow you. Here we aren't normalizing the vector w (we arent' dividing by its norm), instead we are just minimizing its dot product, subject to the constraint. Could you please elucidate a bit more? – Nithish Inpursuit Ofhappiness Mar 14 '17 at 22:12

I'm big on editing posts until they are right, but I have to admit that I'm not quite sure what your question is, and this is just because communication is sometimes hard at both ends. I have some notes on a website that I keep mainly for quick access to info, and I wonder if you could take a look there, and see if we can get a good answer afterwards. – Antoni Parellada Mar 14 '17 at 22:43

Ok, I went though the notes, and my question is related to what comes later in the notes. I understand how each step follows in SVM algorith, and towards the end we are left with this residual problemm - minimize $w^\top w$ subject to a constraint. How is finding a minimal $w^\top w$ relate to the constraint or matter in any way, since we are just scaling down w by minimizing minimize $w^\top w$ – Nithish Inpursuit Ofhappiness Mar 15 '17 at 0:02

@NithishInpursuitOfhappiness Hopefully this is a reasonable answer to your question. – Antoni Parellada Mar 15 '17 at 3:58

1 Notice that we are engaged in an optimization problem (≥ 1), and that this 1, when shifted to the LHS we end up with $y_i \ (x_i \cdot w + b) - 1$, where the b is not defined. There is a great post just dedicated to this point here. – Antoni Parellada Mar 16 '17 at 0:18

1

I think that minimizing the normal vectors, allows you to compare planes just by their normal vectors. If they are the same, so are the two planes. This would not be as trivial if the two normal vectors are scaled differently. There are probably other useful features of having consistent rules for expressing normal vectors which I am not aware of.

Also note that, in terms of the plane, nothing is changing by scaling the normal vector since only the angle (or relative lengths of the components) matters to define a plane. Scaling is simply stretching the normal vector along it's axis.

answered Mar 13 '17 at 22:52