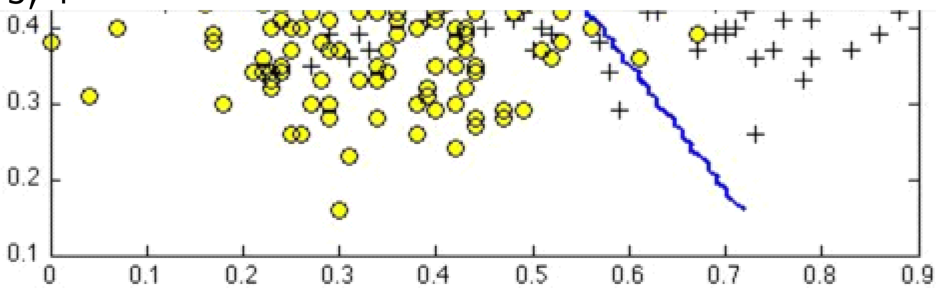
W7-1 Support Vector Machines

Saturday, September 10, 2016 5:55 PM

Right: 1, 2, 3, 4

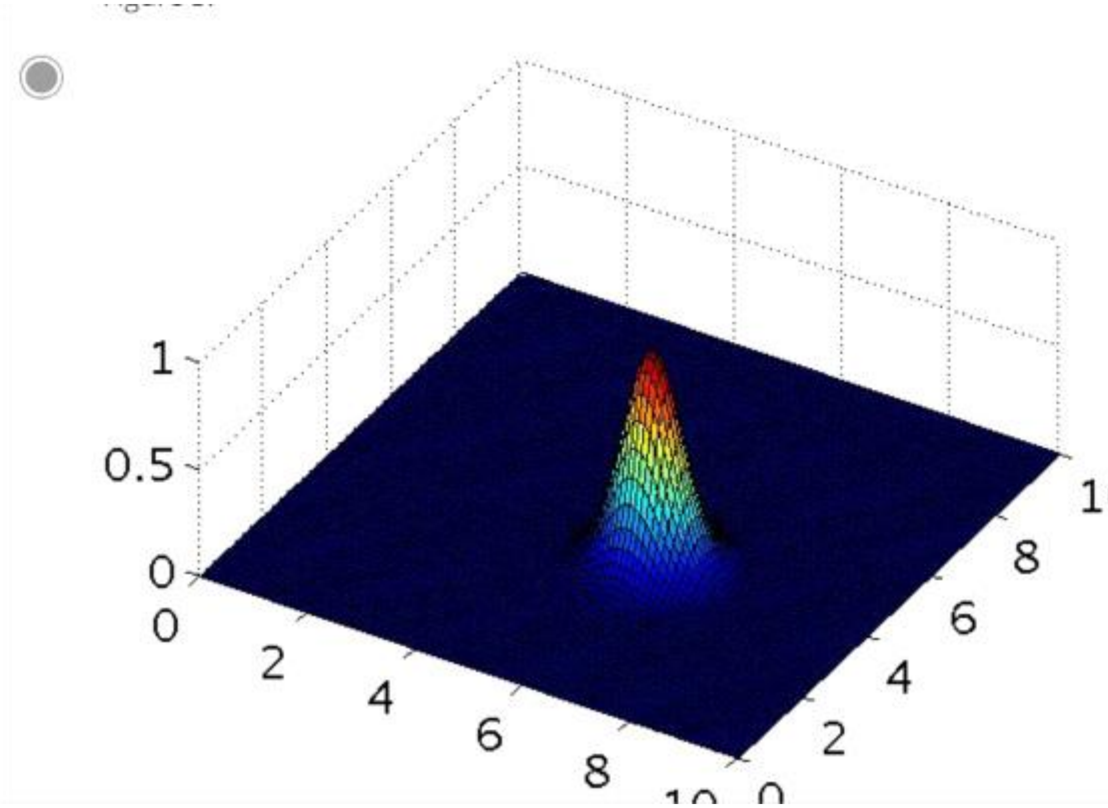


You suspect that the SVM is underfitting your dataset. Should you try increasing or decreasing C? Increasing or decreasing σ^2?

- (●) It would be reasonable to try **increasing** C. It would also be reasonable to try **decreasing** σ^2.
- ( ) It would be reasonable to try **decreasing** C. It would also be reasonable to try **increasing** σ^2.
- ( ) It would be reasonable to try **increasing** C. It would also be reasonable to try **increasing** σ^2.
- ( ) It would be reasonable to try **decreasing** C. It would also be reasonable to try **decreasing** σ^2.

2. The formula for the Gaussian kernel is given by $\text{similarity}(x, l^{(1)}) = \exp\left(-\frac{||x-l^{(1)}||^2}{2\sigma^2}\right)$.

The figure below shows a plot of $f_1 = \text{similarity}(x, l^{(1)})$ when $\sigma^2 = 1$.



The first term in the objective is:

$$C\sum_{i=1}^{m} y^{(i)}\text{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\text{cost}_0(\theta^T x^{(i)}).$$

This first term will be zero if two of the following four conditions hold true. Which are the two conditions that would guarantee that this term equals zero?

- [ ] For every example with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 0$.
- [x] For every example with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 1$.
- [x] For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq -1$.
- [ ] For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq 0$.

4. Suppose you have a dataset with n = 10 features and m = 5000 examples.

After training your logistic regression classifier with gradient descent, you find that it has underfit the training set and does not achieve the desired performance on the training or cross validation sets.

Which of the following might be promising steps to take? Check all that apply.

- [ ] Increase the regularization parameter λ.
- [ ] Use an SVM with a linear kernel, without introducing new features.
- [x] Use an SVM with a Gaussian Kernel.
- [x] Create / add new polynomial features.

5. Which of the following statements are true? Check all that apply.

- [ ] If the data are linearly separable, an SVM using a linear kernel will return the same parameters θ regardless of the chosen value of C (i.e., the resulting value of θ does not depend on C).
- [ ] Suppose you have 2D input examples (ie, $x^{(i)} \in \mathbb{R}^2$). The decision boundary of the SVM (with the linear kernel) is a straight line.
- [ ] The maximum value of the Gaussian kernel (i.e., $sim(x, l^{(1)})$) is 1.
- [x] If you are training multi-class SVMs with the one-vs-all method, it is not possible to use a kernel.