Close

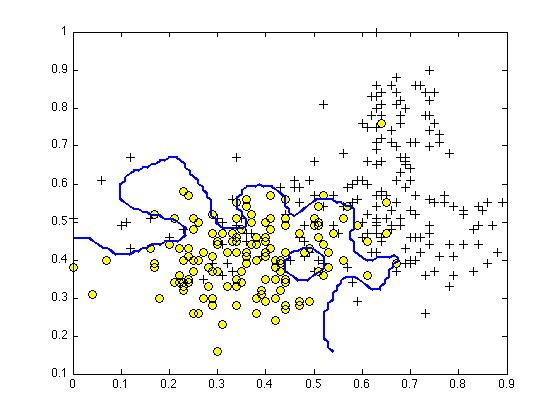
Support Vector Machines

5 questions

1  
point

1.

Suppose you have trained an SVM classifier with a Gaussian kernel, and it learned the following decision boundary on the training set:



When you measure the SVM's performance on a cross validation set, it does poorly. 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**increasing** *σ*2.



It would be reasonable to try **decreasing** *C*. It would also be reasonable to try**decreasing** *σ*2.



It would be reasonable to try **increasing** *C*. It would also be reasonable to try**decreasing** *σ*2.



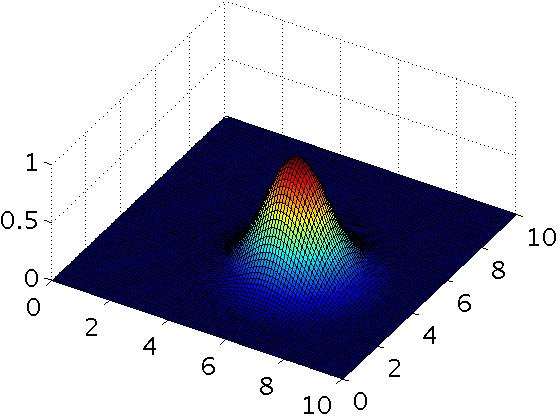
It would be reasonable to try **decreasing** *C*. It would also be reasonable to try**increasing** *σ*2.

1  
point

2.

The formula for the Gaussian kernel is given by similarity(*x*,*l*(1))=exp(−||*x*−*l*(1)||22*σ*2) .

The figure below shows a plot of *f*1=similarity(*x*,*l*(1)) when *σ*2=1.



Which of the following is a plot of *f*1 when *σ*2=0.25?



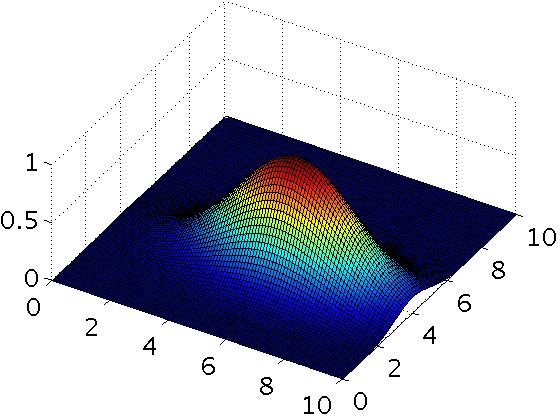


Figure 3.



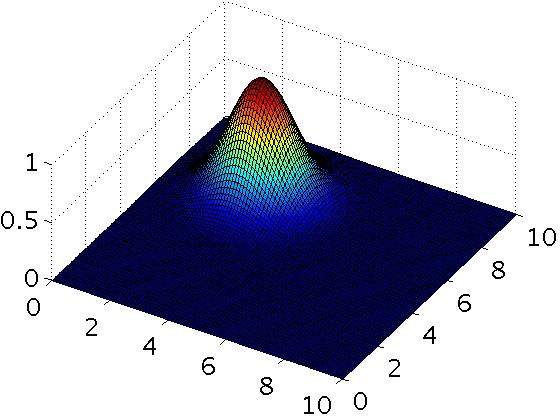


Figure 4.



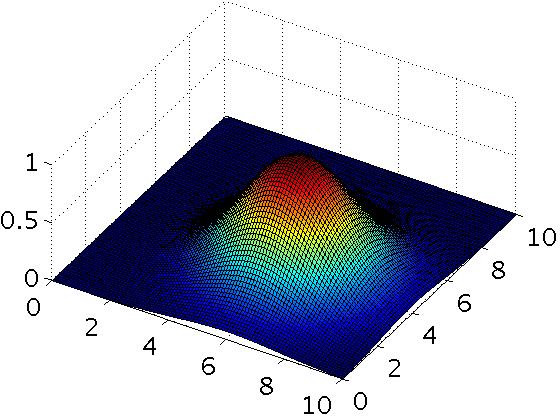
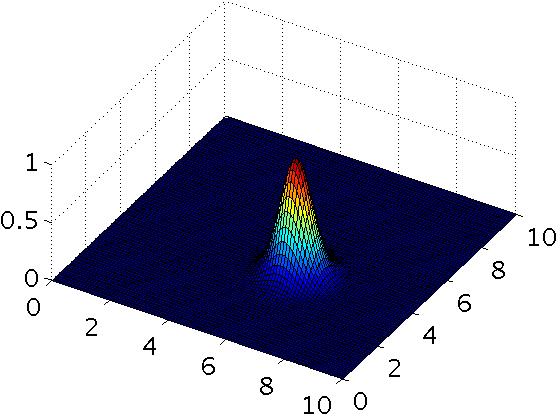


Figure 2.





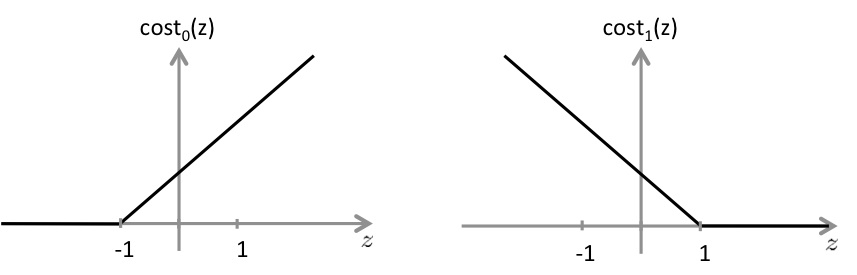
1  
point

3.

The SVM solves

min*θ* *C*∑*mi*=1*y*(*i*)cost1(*θTx*(*i*))+(1−*y*(*i*))cost0(*θTx*(*i*))+∑*nj*=1*θ*2*j*

where the functions cost0(*z*) and cost1(*z*) look like this:



The first term in the objective is:

*C*∑*mi*=1*y*(*i*)cost1(*θTx*(*i*))+(1−*y*(*i*))cost0(*θTx*(*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 *θTx*(*i*)≥0.



For every example with *y*(*i*)=0, we have that *θTx*(*i*)≤0.



For every example with *y*(*i*)=1, we have that *θTx*(*i*)≥1.



For every example with *y*(*i*)=0, we have that *θTx*(*i*)≤−1.

1  
point

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.



Use an SVM with a Gaussian Kernel.



Create / add new polynomial features.



Use an SVM with a linear kernel, without introducing new features.



Increase the regularization parameter *λ*.

1  
point

5.

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



If you are training multi-class SVMs with the one-vs-all method, it is

not possible to use a kernel.



The maximum value of the Gaussian kernel (i.e., *sim*(*x*,*l*(1))) is 1.



Suppose you have 2D input examples (ie, *x*(*i*)∈R2). The decision boundary of the SVM (with the linear kernel) is a straight line.



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*).

2 questions unanswered

Submit Quiz