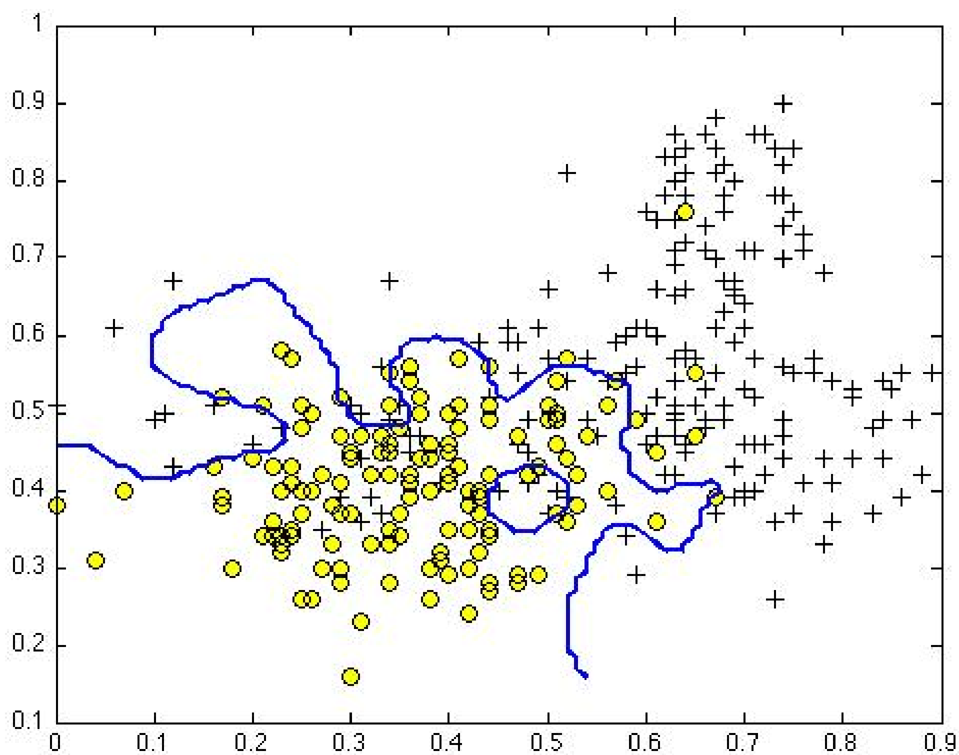
Feedback — XII. Support Vector Machines

Help

You submitted this quiz on **Tue 5 Aug 2014 11:13 AM PDT**. You got a score of **5.00** out of **5.00**.

Question 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?

Your Answer	Score	Explanation
It would be reasonable to try **decreasing** C. It		

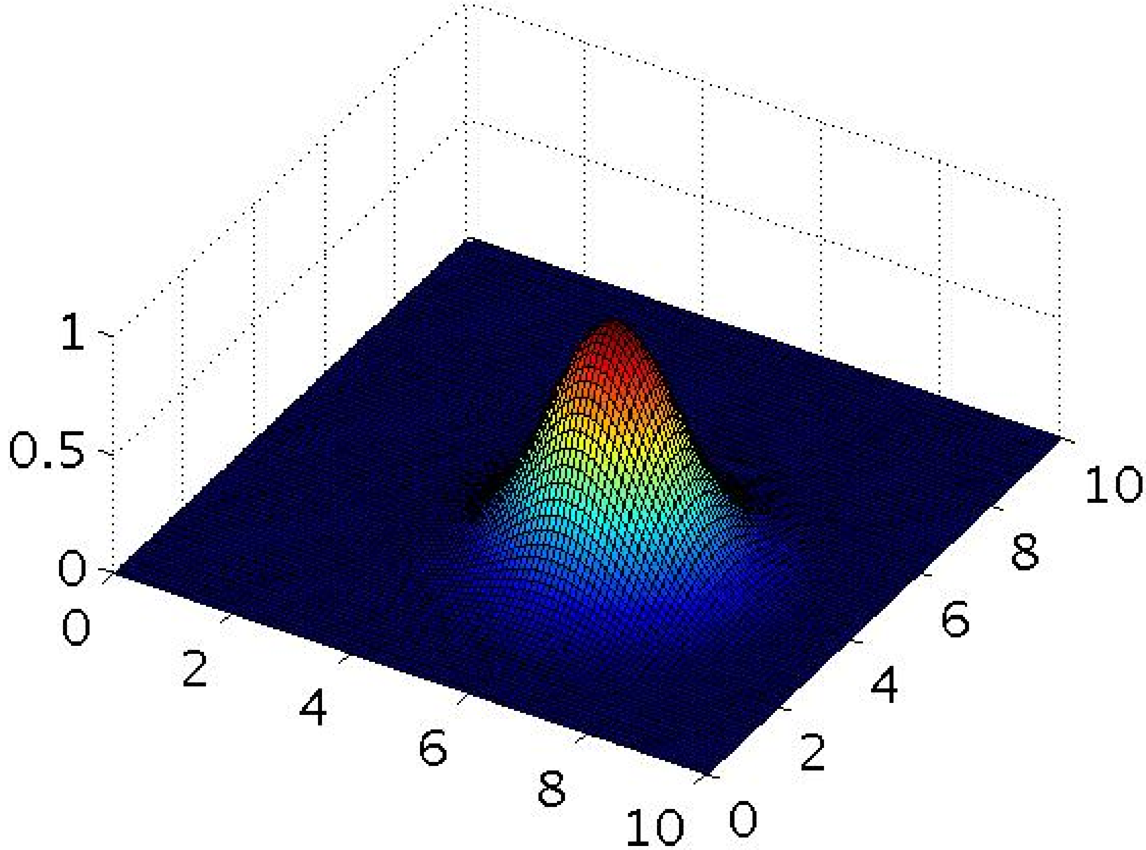
would also be reasonable to try **decreasing** σ^2.

○ It would be reasonable to try **increasing** C. It would also be reasonable to try **increasing** 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.	✔	1.00	The figure shows a decision boundary that is overfit to the training set, so we'd like to increase the bias / lower the variance of the SVM. We can do so by either decreasing the parameter C or increasing σ^2.
Total		1.00 / 1.00	

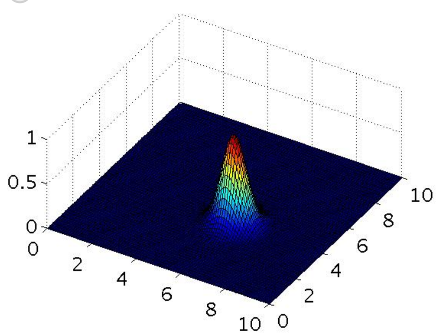
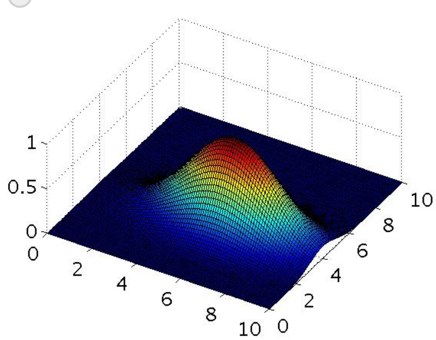
Question 2

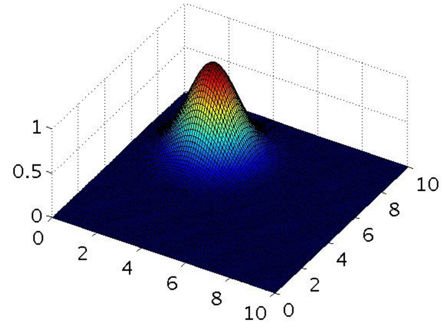
The formula for the Gaussian kernel is given by $\text{similarity}(x, l^{(1)}) = \exp\left(-\frac{||x - l^{(1)}||^2}{2\sigma^2}\right)$.

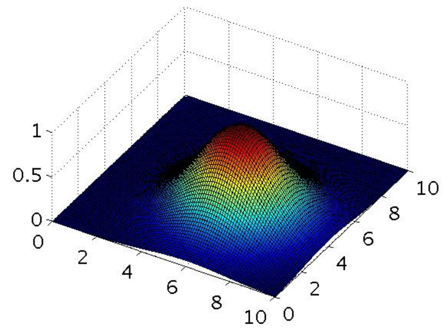
The figure below shows a plot of $f_1 = \text{similarity}(x, l^{(1)})$ when $\sigma^2 = 1$.



Which of the following is a plot of f_1 when $\sigma^2 = 0.25$?

Your Answer		Score	Explanation
	✔	1.00	This figure shows a "narrower" Gaussian kernel centered at the same location which is the effect of decreasing σ^2.
			

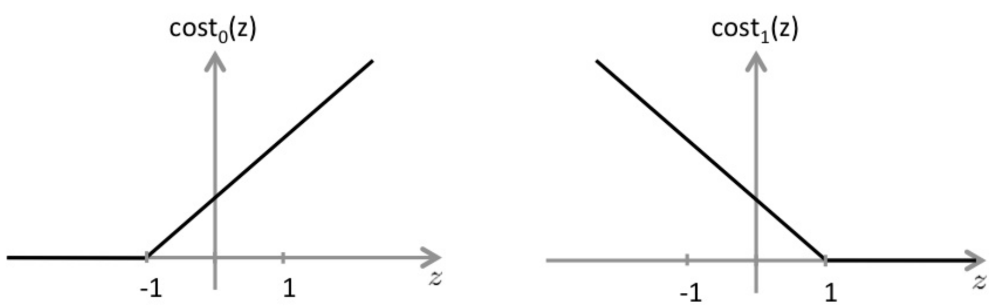




Total	1.00 / 1.00

Question 3

The SVM solves $\min_\theta \; C\sum_{i=1}^m y^{(i)}\text{cost}_1(\theta^T x^{(i)}) + (1-y^{(i)})\text{cost}_0(\theta^T x^{(i)}) + \sum_{j=1}^n \theta_j^2$ where the functions $\text{cost}_0(z)$ and $\text{cost}_1(z)$ look like this:



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?

Your Answer		Score	Explanation
For every example	✔	0.25	$\text{cost}_1(\theta^T x^{(i)})$ is still non-zero for inputs between 0

with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 0$.			and 1, so being greater than or equal to 0 is insufficient.
☑ For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq -1$.	✔	0.25	For examples with $y^{(i)} = 0$, only the $\text{cost}_0(\theta^T x^{(i)})$ term is present. As you can see in the graph, this will be zero for all inputs less than or equal to -1.
☐ For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq 0$.	✔	0.25	$\text{cost}_0(\theta^T x^{(i)})$ is still non-zero for inputs between -1 and 0, so being less than or equal to 0 is insufficient.
☑ For every example with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 1$.	✔	0.25	For examples with $y^{(i)} = 1$, only the $\text{cost}_1(\theta^T x^{(i)})$ term is present. As you can see in the graph, this will be zero for all inputs greater than or equal to 1.
Total		1.00 / 1.00	

Question 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.

Your Answer		Score	Explanation
☑ Use an SVM with a Gaussian Kernel.	✔	0.25	By using a Gaussian kernel, your model will have greater complexity and can avoid underfitting the data.
☐ Reduce the number of examples in the training set.	✔	0.25	While you can improve accuracy on the training set by removing examples, doing so results in a worse model that will not generalize as well.
☑ Create / add new polynomial features.	✔	0.25	When you add more features, you increase the variance of your model, reducing the chances of underfitting.
☐ Use a different optimization method since using gradient descent to	✔	0.25	The logistic regression cost function is convex, so gradient descent will always find the global minimum.

train logistic regression might result in a local minimum.		
Total	1.00 / 1.00	

Question 5

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

Your Answer		Score	Explanation
☐ 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).	✔	0.25	A linearly separable dataset can usually be separated by many different lines. Varying the parameter C will cause the SVM's decision boundary to vary among these possibilities. For example, for a very large value of C, it might learn larger values of θ in order to increase the margin on certain examples.
☐ Suppose you are using SVMs to do multi-class classification and would like to use the one-vs-all approach. If you have K different classes, you will train K - 1 different SVMs.	✔	0.25	The one-vs-all method requires that we have a separate classifier for every class, so you will train K different SVMs.
☑ Suppose you have 2D input	✔	0.25	The SVM without any kernel (ie, the linear kernel) predicts output based only on $\theta^T x$, so it gives a linear / straight-line

examples (ie, $x^{(i)} \in \mathbb{R}^2$). The decision boundary of the SVM (with the linear kernel) is a straight line.			decision boundary, just as logistic regression does.
It is important to perform feature normalization before using the Gaussian kernel.	✔	0.25	The similarity measure used by the Gaussian kernel expects that the data lie in approximately the same range.
Total		1.00 / 1.00	