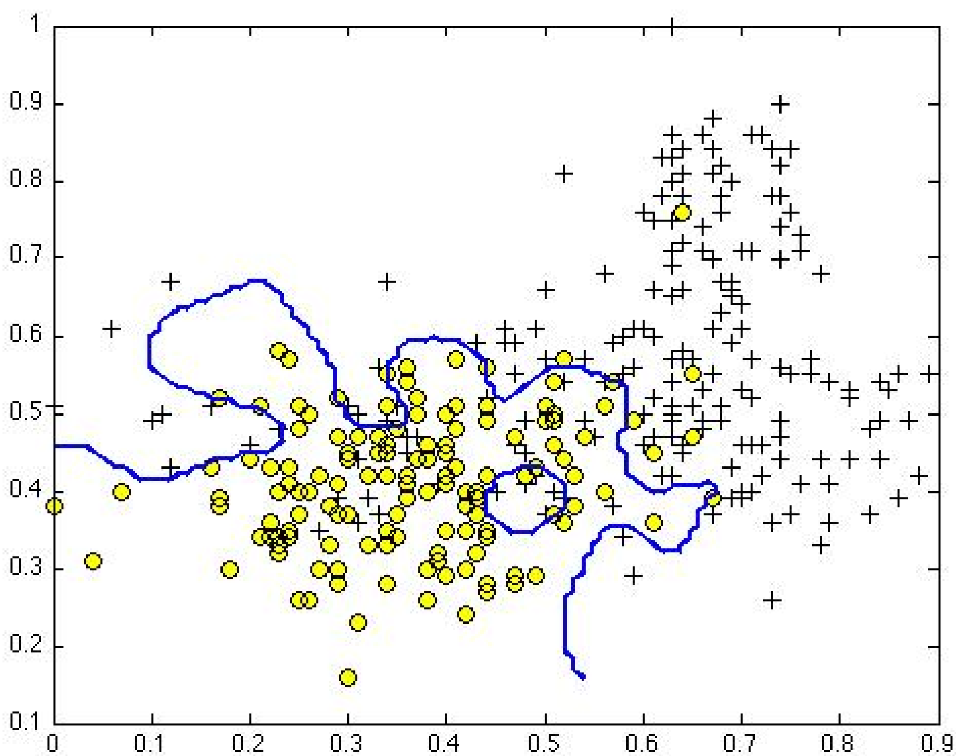
Feedback — XII. Support Vector Machines

Help

You submitted this quiz on **Sun 15 Dec 2013 9:59 PM PST**. 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:



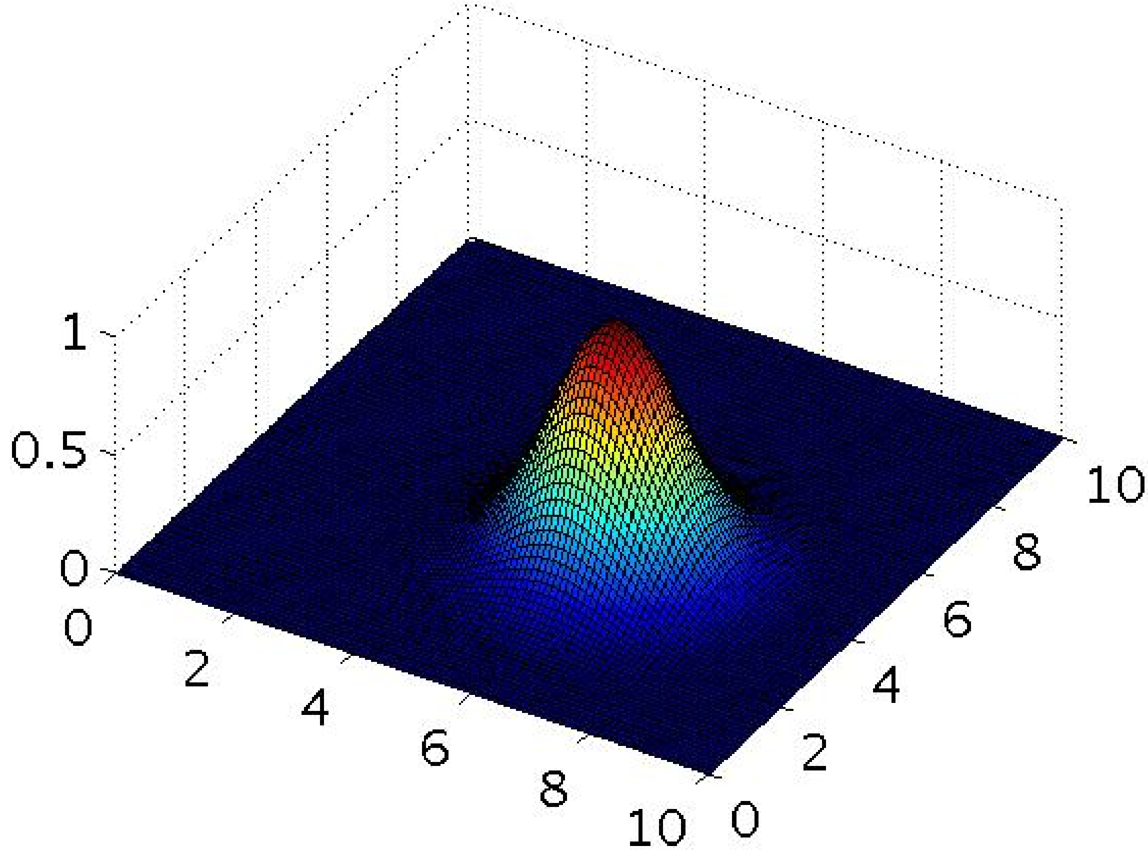
When you measure the SVM's performance on a cross validation set, it does poorly. Should you try increasing or decreasing C? Increasing or decreasing σ^2?

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

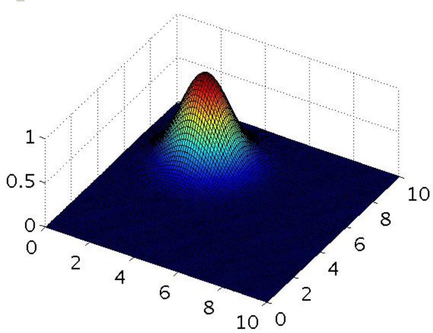
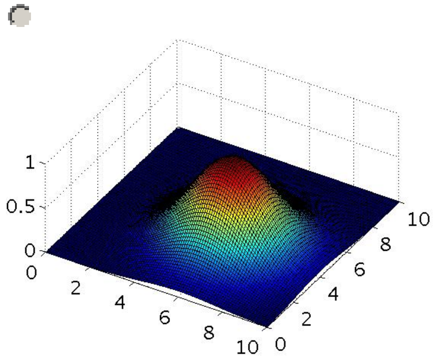
reasonable to try **decreasing** σ^2.			
It would be reasonable to try **decreasing** C. It would also be reasonable to try **decreasing** σ^2.			
It would be reasonable to try **increasing** C. It would also be reasonable to try **increasing** σ^2.			
It would be reasonable to try **decreasing** C.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

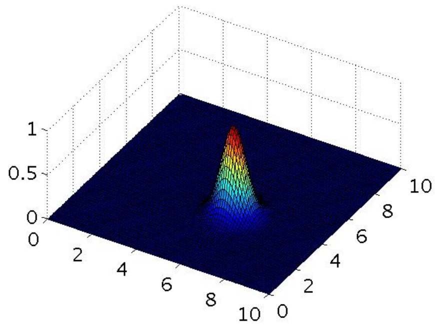
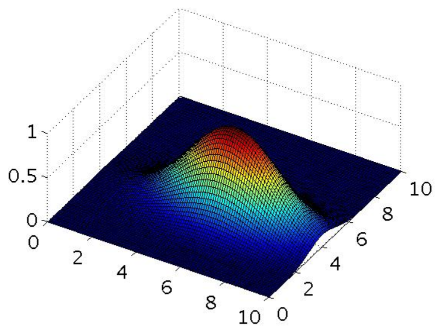
Question 2

The formula for the Gaussian kernel is given by $\mathrm{similarity}(x, l^{(1)}) = \exp\left(-\frac{||x-l^{(1)}||^2}{2\sigma^2}\right)$. The figure below shows a plot of $f_1 = \mathrm{similarity}(x, l^{(1)})$when $\sigma^2 = 1$.



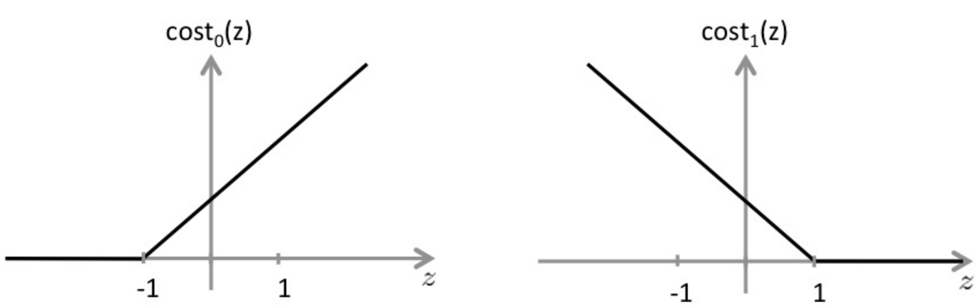
Which of the following is a plot of f_1 when $\sigma^2 = 0.25$?

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 with $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 1$.	✔	0.25	For examples with $y^{(i)} = 1$, only the $\mathrm{cost}_1(\theta^T x^{(i)})$ term is present. As you can see in the graph, this will be zero for all inputs greater than or equal to 1.
☐ For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq 0$.	✔	0.25	$\mathrm{cost}_0(\theta^T x^{(i)})$ is still non-zero for inputs between -1 and 0, so being less than or equal to 0 is insufficient.
☑ For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq -1$.	✔	0.25	For examples with $y^{(i)} = 0$, only the $\mathrm{cost}_0(\theta^T x^{(i)})$ 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 $y^{(i)} = 1$, we have that $\theta^T x^{(i)} \geq 0$.	✔	0.25	$\mathrm{cost}_1(\theta^T x^{(i)})$ is still non-zero for inputs between 0 and 1, so being greater than or equal to 0 is insufficient.
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 linear kernel, without introducing new features.	✔	0.25	An SVM with only the linear kernel is comparable to logistic regression, so it will likely underfit the data as well.
☑ Try using a neural network with a large number of hidden units.	✔	0.25	A neural network with many hidden units is a more complex (higher variance) model than logistic regression, so it is less likely to underfit the data.

☐ Increase the regularization parameter λ.	✔	0.25	You are already underfitting the data, and increasing the regularization parameter only makes underfitting stronger.
☑ Create / add new polynomial features.	✔	0.25	When you add more features, you increase the variance of your model, reducing the chances of underfitting.
Total		1.00 / 1.00	

Question 5

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

Your Answer		Score	Explanation
☐ If you are training multi-class SVMs with the one-vs-all method, it is not possible to use a kernel.	✔	0.25	Each SVM you train in the one-vs-all method is a standard SVM, so you are free to use a kernel.
☑ 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.	✔	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 decision boundary, just as logistic regression does.
☐ 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.
☑ The maximum value of the Gaussian kernel (i.e., $sim(x, l^{(1)})$) is 1.	✔	0.25	When $x = l^{(1)}$, the Gaussian kernel has value $\exp(0) = 1$, and it is less than 1 otherwise.

Total	1.00 / 1.00