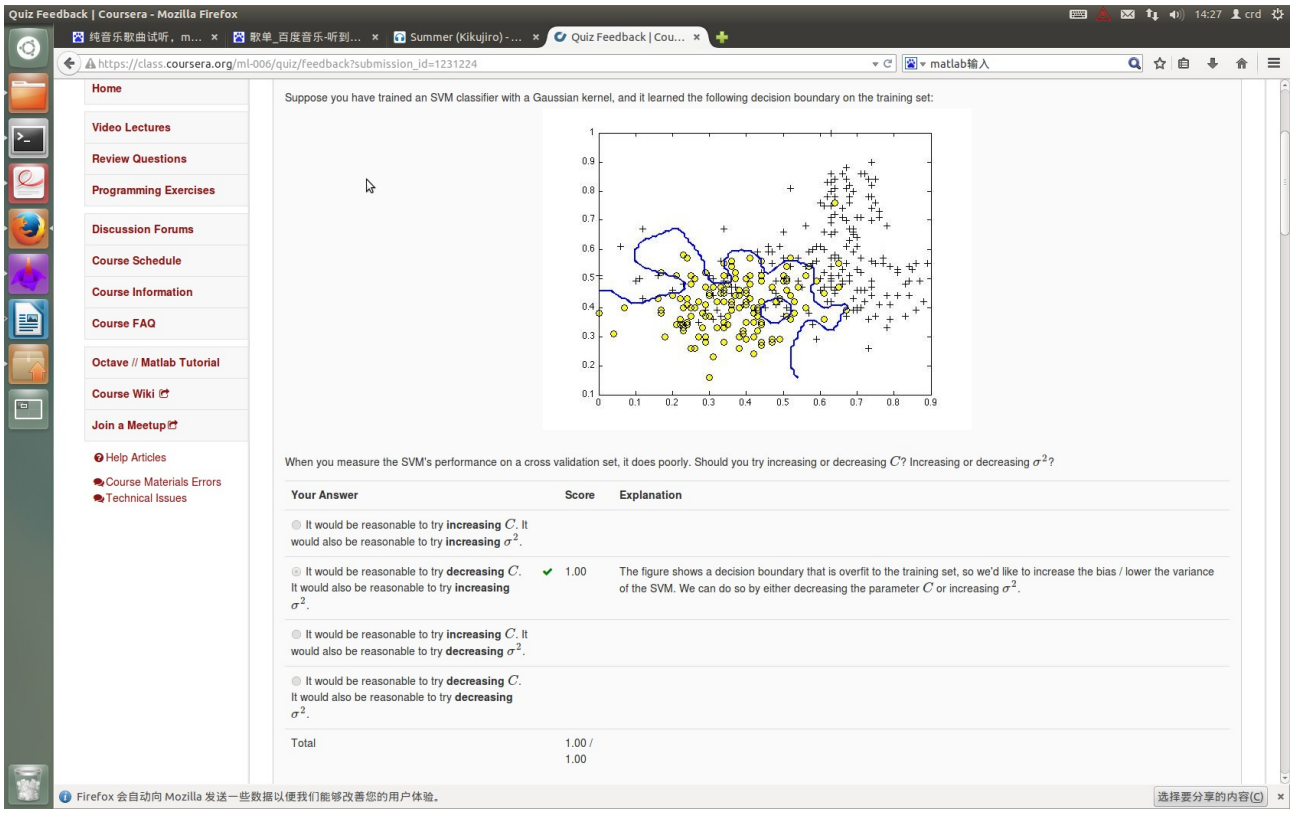
m 个 training example，L 是网络层数，sl 是某一层的节点个数
正则项没有加入 theta0，J 是 cost function，神经网络则对每一个输出节点合成了 J
bp 算法，求 J 对 theta 的偏导，前面的都是 g(),最后一层是 h(),delta 表达的是每一层的 error，（a-y），其中 a 就是 h()，没有 detla1
将差错传递回来，
theta 的上标代表第几层，脚标代表是第几个神经元（j）到下一层的哪个神经元（i），表达式为 ij，下一层在前，上一层在后
与前向传播的表达式很像
可以将 logsitic 造成的 cost function 类比于平方差
delta 是 cost 关于 z 的偏导项
假如 z 有改变，那么会改变 h(x),从而改变 cost
因此 detla 由后面一层的 delat 乘以 weight 得到
【unrolling】将参数展开成向量，神经网络中的不是向量，
octave 里面，这样就将 theta1，2，3 里面的元素全部取出来展开成一个长向量
将向量再重新展开为矩阵
因为写 costfunction 代码时，需要传入的是一个长向量
fminunc 进行优化的时候需要将初始参数长向量传递进去
【gradient checking】保证求导正确
首先数值上估计导数，可以选取两边的两个点连线的斜率近似导数
当有多个 theta，也就是将每个 theta 的求导单独考虑，epsilon 相同，再对比 bp 算出的导数
【初始化参数】initialTheta，假如全部初始化为 0，那么就只有 bias 在相加，那么会导致计算出的 weight 是相同的
上图相同颜色的线就一直相同，那么计算出来的特征就都是相同的，初始化的话需要相同大小的正负边界。



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. |
| Increase the regularization parameter $\lambda$. | ✔ | 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. |
| 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. |
| Total | | 1.00 / 1.00 | |

## Question 5

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

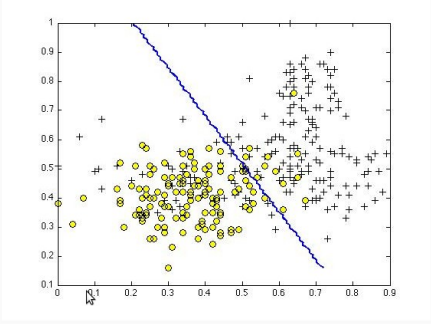
| Your Answer | | Score | Explanation |
|---|---|---|---|
| Suppose you have 2D input examples (ie, $x^{(i)} \in \mathbb{R}^2$). The decision boundary of the SVM (with the linear kernel) is a straight line. | ✔ | 0.25 | The SVM without any kernel (ie, the linear kernel) predicts output based only on $\theta^T x$, 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. |
| If you are training multi-class SVMs with the one-vs-all method, it is not possible to use a kernel. | ✖ | 0.00 | Each SVM you train in the one-vs-all method is a standard SVM, so you are free to use a kernel. |
| The maximum value of the Gaussian kernel (i.e., $sim(x, l^{(1)})$) is 1. | ✖ | 0.00 | When $x = l^{(1)}$, the Gaussian kernel has value $\exp(0) = 1$, and it is less than 1 otherwise. |
| Total | | 0.50 / 1.00 | |





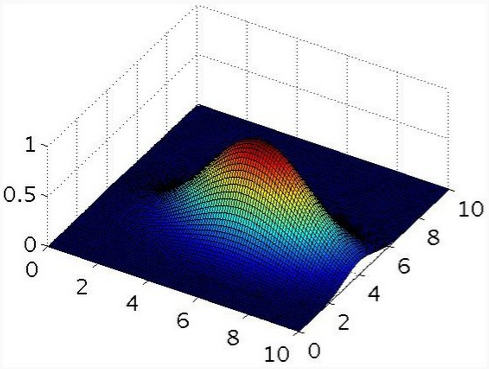
Home
Video Lectures
Review Questions
Programming Exercises
Discussion Forums
Course Schedule
Course Information
Course FAQ
Octave // Matlab Tutorial
Course Wiki
Join a Meetup
Help Articles
Course Materials Errors
Technical Issues

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




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

| Your Answer | | Score | Explanation |
|---|---|---|---|
| It would be reasonable to try **increasing** $C$. It would also be reasonable to try **decreasing** $\sigma^2$. | ✔ | 1.00 | The figure shows a decision boundary that is underfit to the training set, so we'd like to lower the bias / increase the variance of the SVM. We can do so by either increasing the parameter $C$ or decreasing $\sigma^2$. |
| It would be reasonable to try **decreasing** $C$. It would also be reasonable to try **increasing** $\sigma^2$. | | | |
| It would be reasonable to try **decreasing** $C$. It would also be reasonable to try **decreasing** $\sigma^2$. | | | |
| It would be reasonable to try **increasing** $C$. It would also be reasonable to try **increasing** $\sigma^2$. | | | |
| Total | | 1.00 / 1.00 | |



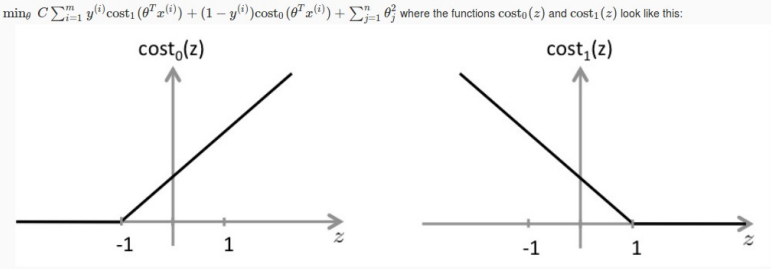

✖ 0.00 This figure shows an ellipcial Gaussian kernel which cannot occur when there is a single variance parameter $\sigma^2$.

Quiz Feedback | Coursera - Mozilla Firefox

https://class.coursera.org/ml-006/quiz/feedback?submission_id=1231478

## 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 $\text{cost}_1(\theta^T x^{(i)})$ term is present. As you can see in the graph, this will be zero for all inputs greater than or equal to 1. |
| For every example with $y^{(i)} = 0$, we have that $\theta^T x^{(i)} \leq -1$. | ✖ 0.00 | 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 $\theta^T x^{(i)} \geq 0$. | ✔ 0.25 | $\text{cost}_1(\theta^T x^{(i)})$ is still non-zero for inputs between 0 and 1, so being greater than or equal to 0 is insufficient. |
| Total | 0.75 / 1.00 | |

Firefox 会自动向 Mozilla 发送一些数据以便我们能够改善您的用户体验。