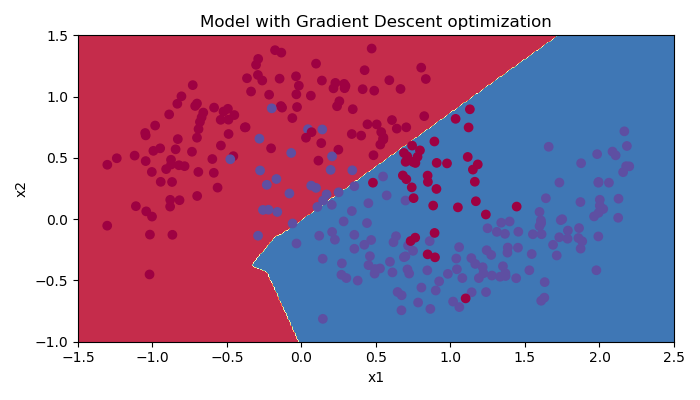
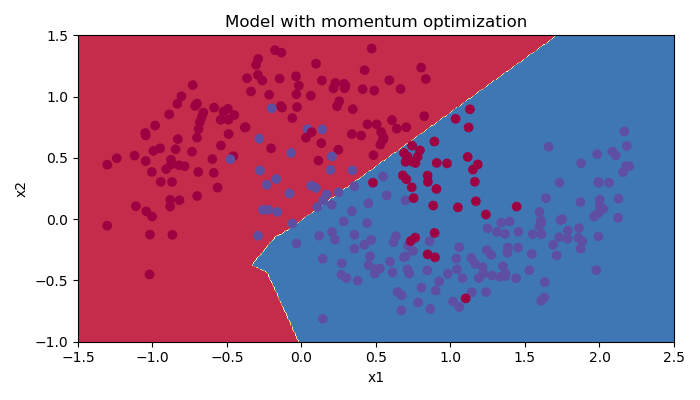
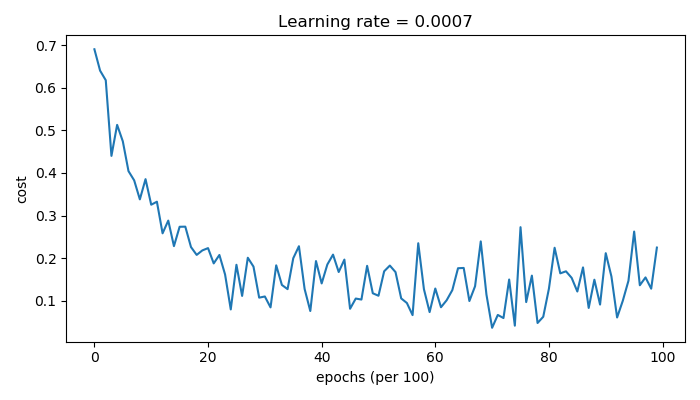
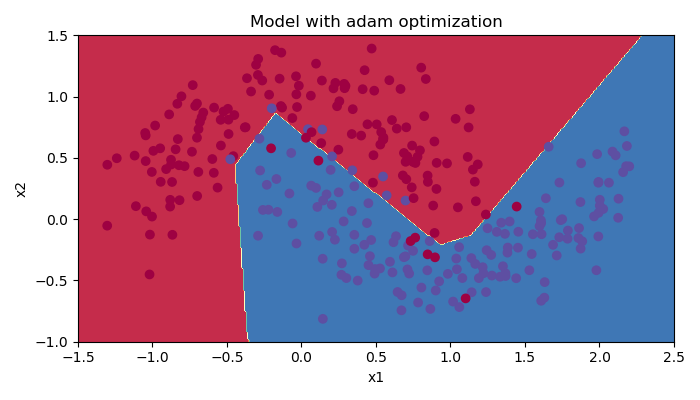
## 梯度下降，accuracy：0.796666

## 带有动量的 accuracy：0.796666



## Adam：accuracy：0.943333



## 总结：

具有动量的梯度下降通常可以有很好的效果，但由于小的学习速率和简单的数据集所以它的影响几乎是轻微的。另一方面，Adam明显优于小批量梯度下降和具有动量的梯度下降，如果在这个简单的模型上运行更多时间的数据集，这三种方法都会产生非常好的结果，然而，我们已经看到Adam收敛得更快。

  Adam的一些优点包括相对较低的内存要求（虽然比梯度下降和动量下降更高）和通常运作良好，即使对参数进行微调（除了学习率α

）

