1. **Gradients of some basic layers:**
   1. **BatchNorm layer**

 和 为尺度变化系数，求其偏导：





由于后面的计算需要使用批归一化中的，所以我们在本节进行推导。

先由 推导和：





然后求出，和：







最后求出：



* 1. **dropout layer**

将第i个输入节点表示为，将输出节点表示为，将mask表示为。 然后，相对于droupout层的输入，输出的梯度是：



* 1. **softmax**

对于softmax激活函数，输入 ，输出



softmax激活函数的第i个输出相对于第j个输入的梯度为：



整理得，其中为指示函数，时为1，时为0：



1. **Feed-forward and backpropagation of the predefined multi-task network**
   1. **Feed forward computation**

首先推导任务A的前向计算， FC1A 层的输出是:

 

DP1A 层的输出是:

 

FC1A 层也就是任务A的输出是:

 

然后推导任务B的前向计算， FC1B 层的输出是:

 

BN1B层的输出是:

 

FC2B 层也就是任务B的输出是:

 

然后推导任务C的前向计算， FC1C 层的输出是:

 

BN1C层的输出是:

 

FC2C 层也就是任务C的输出是:

 

* 1. **Backpropagation**

**首先推导任务C的反向传播**。任务C使用交叉熵损失函数，对于FC2C 层，其中n为类别数，字母上标代表维度序号（类别）:







对于BN1C 层，1.1节推导出的公式可以直接套用:













对于FC1C 层：





**然后推导任务B的反向传播**。任务B使用平方误差损失，字母上标代表维度序号，对于 ，梯度是：



对于FC2B 层:





对于BN1B 层，1.1节推导出的公式可以直接套用:













对于FC1B 层：





**然后推导任务A的反向传播**。任务A使用平方误差损失，字母上标代表维度序号，对于 ，梯度是：



对于FC2A 层:





对于DP1A 层，1.2节推导出的公式可以直接套用:





对于FC1A 层：



