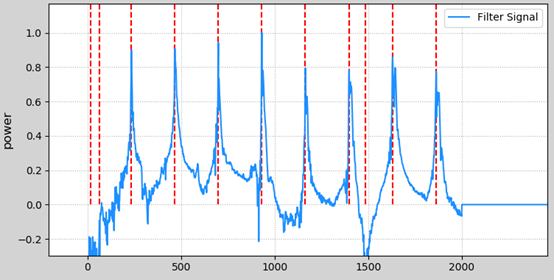
基于频域做的特征工程



模型一：训练数据 3D张量格式：对时域长度为0.1s的音频序列没做任何变换直接输入[样本数量，时长，幅度]，标签为对应的频率

深度网络架构:LSTM+Dense+回归模型

在2407个样本中正确率0.149

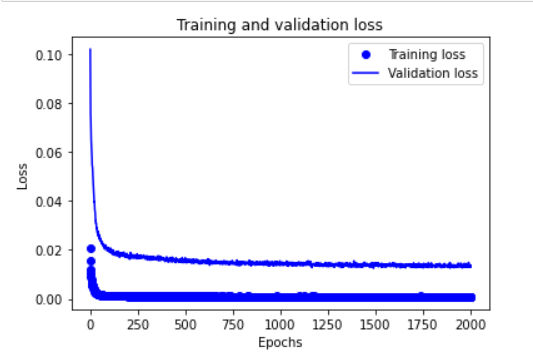
模型二：训练数据 2D张量格式 [样本数，[特征峰，幅度]] 训练标签：音高所对应的频率

采用模型：三层Dense+回归模型

在2407个样本中正确率0.574

模型三：训练数据 3D张量格式 [样本数，频率，幅度] 训练标签：音高所对应的频率

深度网络架构:LSTM+Dense+回归模型



在2407个样本中正确率0.514我们思考可能对于机器学习的回归模型不断优化的是预测出频率，而不是音高，所以可能分类模型更加适合。另外，采用LSTM网络的时候网络过于复杂，参数过多，训练的效率较低。

模型四：最终方案

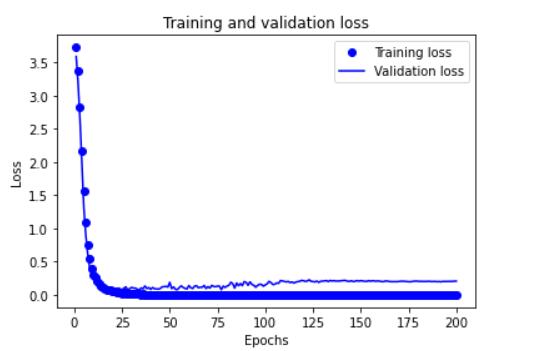
使用傅里叶变换后的频域数据输入网络并调试网络参数，成功将预测率提升到94%

训练样本：傅里叶变换后的[频率，幅值]序列，样本标签：对应的音高类别

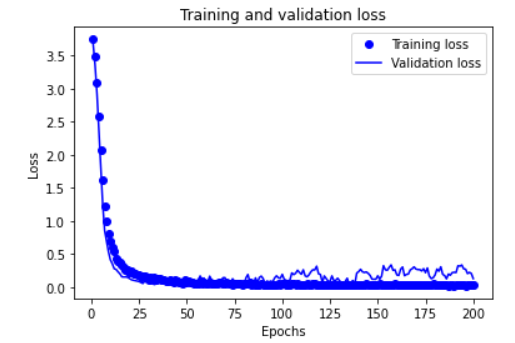
网络模型：CNN网络+多分类模型

使用dropout方法：随机断开全连接层的连接 对网络进行优化 出现了轻度过拟合，在真实数据集的表现差于训练集的表现

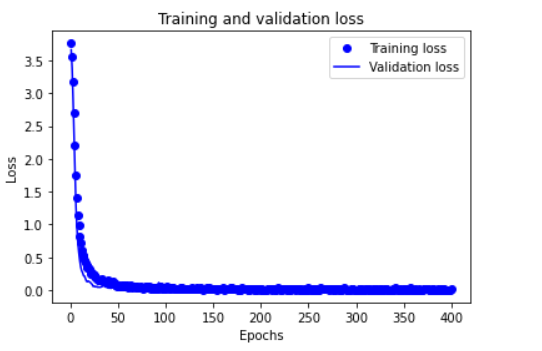
丢失率=0



丢失率=0.10



丢失率=0.13



**当丢失率超过0.2时，准确率直接降至50%一下，最终采取dropout=0.13**

**最后在2407个样本中：准确率由94%上升至97.2%**