1. 网络结构：

网络结构是一个三层的MLP，包括输入层、隐藏层和输出层。

每一层都由线性层和ReLU激活函数组成，用于引入非线性性质和增强网络的表达能力。

输入层的维度为784（对应图像的像素数），隐藏层的维度为512，输出层的维度为10（对应10个类别）。

1. 训练过程：

训练过程包括对网络参数的迭代更新，以最小化损失函数。

训练集被随机分为小批量进行训练，每个批量包含一定数量的图像样本和对应的标签。

对于每个批量，网络首先进行前向传播计算预测结果，然后计算预测结果与真实标签之间的损失。

损失通过反向传播算法传递回网络，根据梯度信息更新网络参数。

训练过程通过多个epoch（迭代次数）进行，每个epoch中都对整个训练集进行一次完整的训练。

1. 优化器选择：

优化器用于更新网络参数，以最小化损失函数。

在示例代码中，使用了随机梯度下降（SGD）优化器，它通过计算样本的梯度来更新参数。

SGD优化器使用一个固定的学习率（lr）来控制参数更新的步长，同时利用动量（momentum）来加速梯度下降过程。