用Python进行机器学习开发

**一、主讲老师**

顾翔

**二、培训内容背景**

机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。它是人工智能核心，是使计算机具有智能的根本途径。

**三、课程介绍**

本课程介绍机器学习的python实现。包括“数据分析”“机器学习算法”“数据处理和优化”三大部分。数据分析中主要介绍“基本库”和“数据分析”。机器学习算法中主要介绍“原理”“K邻近算法(KNeighbors)”“朴素贝叶斯”“支持向量机(SVM)”“决策树(DecisionTree)”“集成学习”“聚类”“降维”“神经网络(MLP)”的理念和如何用Python实现。数据处理和优化介绍”数据处理“”数据表达与特征工程“”模型评估“和”管道模型“。

**四、培训对象**

　机器学习开发、测试、运维工程师

**五、学员基础**

　熟悉掌握Python语言、有一定的高等数学、线性代数和概率论知识

**六、教学时间**

5天

**七、课程内容**

**1数据分析**

**1.1基本库**

**Numpy**

**Pandas**

**Matplotlib**

**Scipy**

**1.2数据分析**

**数据加载**

**cvs文件**

**Excel文件**

**数据库**

**数据清洗和准备**

**数据概览和类型转换**

**处理丢失数据**

**处理重复数据**

**数据转换**

**数据替换**

**数据离散化**

**数据拆分**

**过滤异常值**

**字符串处理**

**数据规整**

**层次化索引**

**合并数据集**

**重塑和轴向旋转**

**可视化**

**折线**

**柱状图**

**直方图和密度图**

**散点图**

**分组聚合**

**分组**

**聚合**

**基本聚合**

**桶分析**

**2 机器学习算法**

**原理**

**线性模型(Linear)**

**线性模型原理**

**线性回归(LinearRegression)**

**StatsModels的线性回归**

**逻辑回归算法(Logistical Regression)**

**岭回归算法(Ridge Regression)**

**套索回归算法(Lasso Regression)**

**弹性网络(Elastic Net)**

**K邻近算法(KNeighbors)**

**K邻近算法原理**

**K邻近分类算法(KNeighborsClassifier)**

**K邻近回归算法(KNeighborsRegressor)**

**朴素贝叶斯**

**朴素贝叶斯原理(NB)**

**贝努利贝叶斯(BernoulliNB)**

**高斯贝叶斯(GaussianNB)**

**多项式贝叶斯(MultinomialNB)**

**支持向量机(SVM)**

**支持向量机原理**

**支持向量机分类算法(SVC)**

**支持向量机线性分类算法(LinearSVC)**

**支持向量机回归算法(SVR)**

**支持向量机线性回归算法(LinearSVR)**

**决策树(DecisionTree)**

**决策树原理**

**决策树分类算法(DecisionTreeClassifier)**

**决策树回归算法(DecisionTreeRegressor)**

**集成学习**

**集成学习原理**

**随机森林算法(RandomForest)**

**随机森林分类算法(RandomForestClassifier)**

**随机森林回归算法(RandomForestRegressor)**

**AdaBost(Adaptive Boosting)**

**装袋算法(Bagging)**

**投票分类(Voting Classifier)**

**堆垛分类(Stacking Classifier)**

**聚类**

**聚类原理**

**K均值聚类(k-means)**

**凝聚聚类(agglomerative)**

**DBSCAN**

**降维**

**降维原理**

**主生成分析(PCA)**

**非负矩阵分解(NMF)**

**线性判别分析(LDA)**

**神经网络(MLP)**

**神经网络原理**

**神经网络分类算法(MLPClassifier)**

**神经网络分类算法(MLPRegressor)**

**3 数据处理和优化**

**数据处理**

**数据表达与特征工程**

**模型评估**

**管道模型**