

# CLASSIFICAÇÃO E ANÁLISE DE DESEMPENHO DE FUNDOS

Aluno: José Corsini Filho - Orientador: Prof. Dr. Renato Vicente  
Instituto de Matemática e Estatística - USP

## Introdução

Um fundo de investimento é uma ferramenta financeira de aplicação que reúne recursos de um conjunto de investidores, visando a obtenção de ganhos financeiros a partir de uma carteira de investimentos. A estratégia adotada e o segmento em que um fundo atua podem ser representados pelas classificações (categorias), divulgadas por diversas fontes que atuam no mercado financeiro (públicas e privadas). Estas classificações são extremamente úteis para a análise do desempenho de um fundo em relação à sua categoria e outros segmentos de investimento.

A construção das classificações envolve uma combinação entre conhecimento técnico e criatividade. A obtenção de agrupamentos condizentes com o cenário desejado para análise do usuário final (gestor do fundo, investidor, dentre outros) será consequência das técnicas de agrupamento, indicadores (qualitativos e quantitativos) e parametrizações utilizados.

O objetivo deste trabalho é realizar a implementação de uma ferramenta que colete dados históricos dos fundos de investimento do mercado e realize o processo de classificação (*clusterização*), disponibilizando algumas estratégias estatísticas já conhecidas, como o *k-means*, por exemplo. Além de utilizar e integrar diversas tecnologias, uma das preocupações no ambiente de desenvolvimento foi proporcionar flexibilidade para a inclusão de novos indicadores e algoritmos de agrupamento. A quantidade de fundos disponíveis no mercado e a minimização do número de *outliers* nas classificações são algumas das motivações que se tornam desafios nesta proposta.

## Classificação (*Clusterização*)

Inicialmente, foram estudados uma dissertação de mestrado [Sil07] e um trabalho de conclusão de curso [Cas10], os quais nortearam a escolha e o entendimento do comportamento de alguns indicadores (como o *Beta* do modelo CAPM, por exemplo), além de alguns algoritmos para a classificação dos fundos, dentre eles o *k-means*.

O método *k-means* é uma técnica de agrupamento de dados [Eve11] apresentada por James MacQueen em 1967. Dada uma amostra de elementos a serem analisados, são definidos heurísticamente *k-means* centróides como base para o processo. De acordo com uma medida de similaridade definida, os elementos da amostra são alocados em grupos associados aos centróides. Como próximo passo, os centróides de cada grupo são recalculados e é realizada uma nova iteração do processo de enquadramento. O processo é encerrado em 2 circunstâncias: uma nova iteração não surte realocações de elementos ou ao se atingir o número limite de iterações pré-definido.

Na FIGURA 1, ilustramos um exemplo bi-dimensional da aplicação do *k-means* em uma amostra com 1000 elementos, gerados aleatoriamente através do programa R.

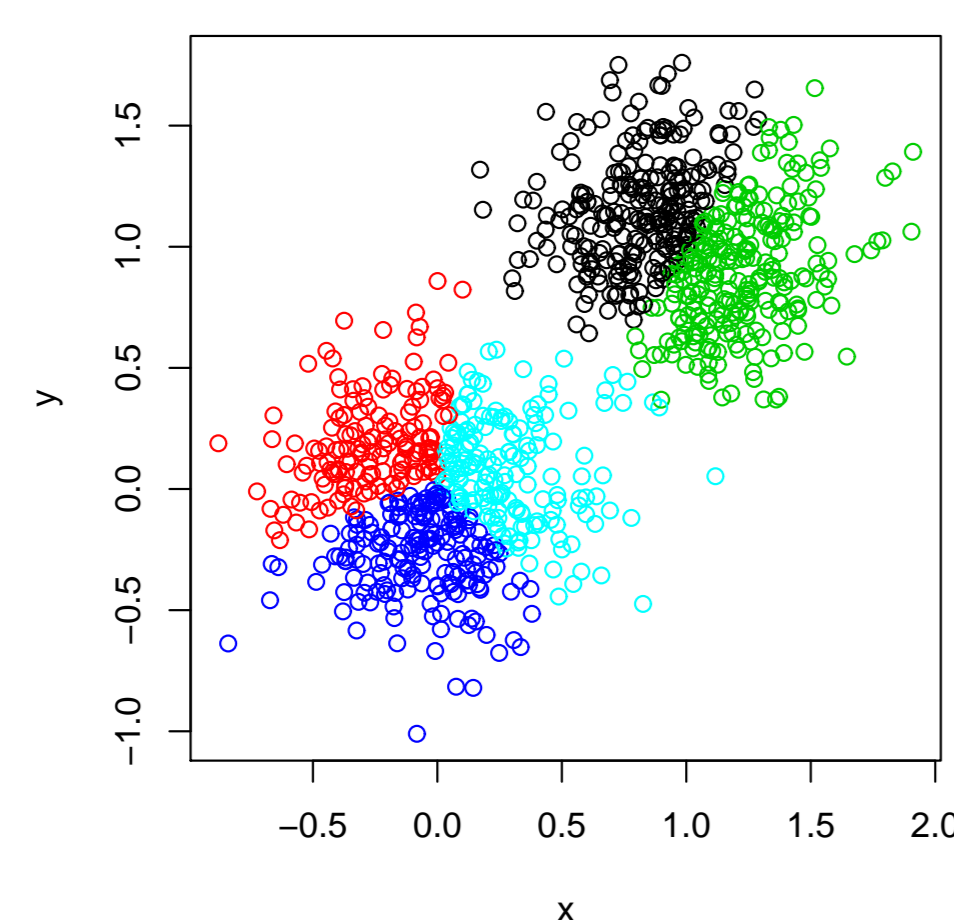


FIGURA 1: *k-means*

## Análise de Categoria

Após o processo de classificação ser realizado, torna-se interessante examinar as categorias obtidas em particular, buscando melhorar a qualidade de dados dentro de cada grupo. Para analisar o desempenho dos fundos de uma determinada categoria, propomos aqui um simples estudo com o gráfico de dispersão entre risco e retorno dos mesmos. Notamos que, mesmo após a utilização do algoritmo de agrupamento, uma categoria ainda pode apresentar *outliers*, isto é, observações numericamente distantes da tendência do grupo, as quais podem comprometer significativamente determinadas análises comparativas. A FIGURA 2 apresenta um exemplo de um gráfico de dispersão (risco x retorno) para uma determinada categoria obtida:

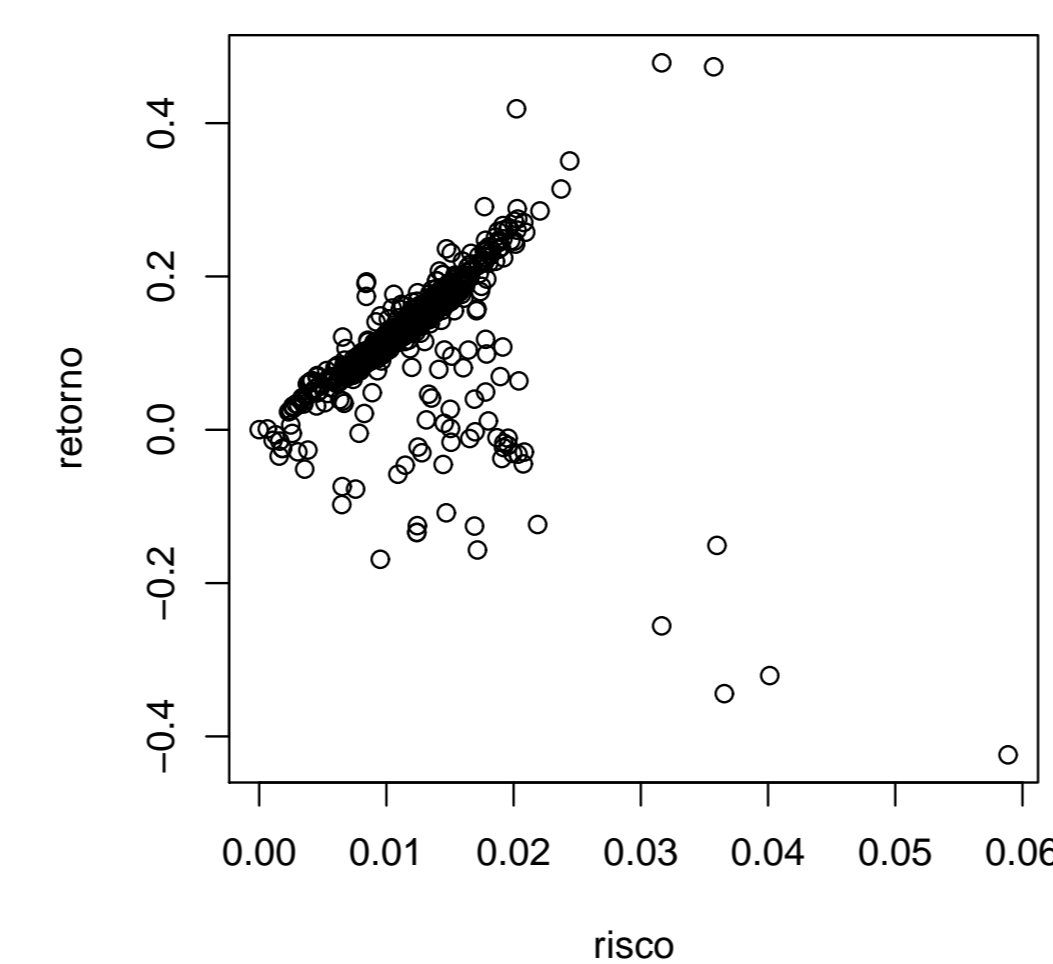


FIGURA 2: Dispersão após classificação

Uma proposta para a solução deste problema é a utilização da *Distância de Mahalanobis* [Had92], uma medida de distância que leva em consideração as correlações entre os elementos estudados. Assim, é possível detectar estes *outliers* e removê-los da amostra inicial, obtendo uma melhor concentração dos dados. A FIGURA 3 ilustra a aplicação desta técnica com base na amostra anterior:

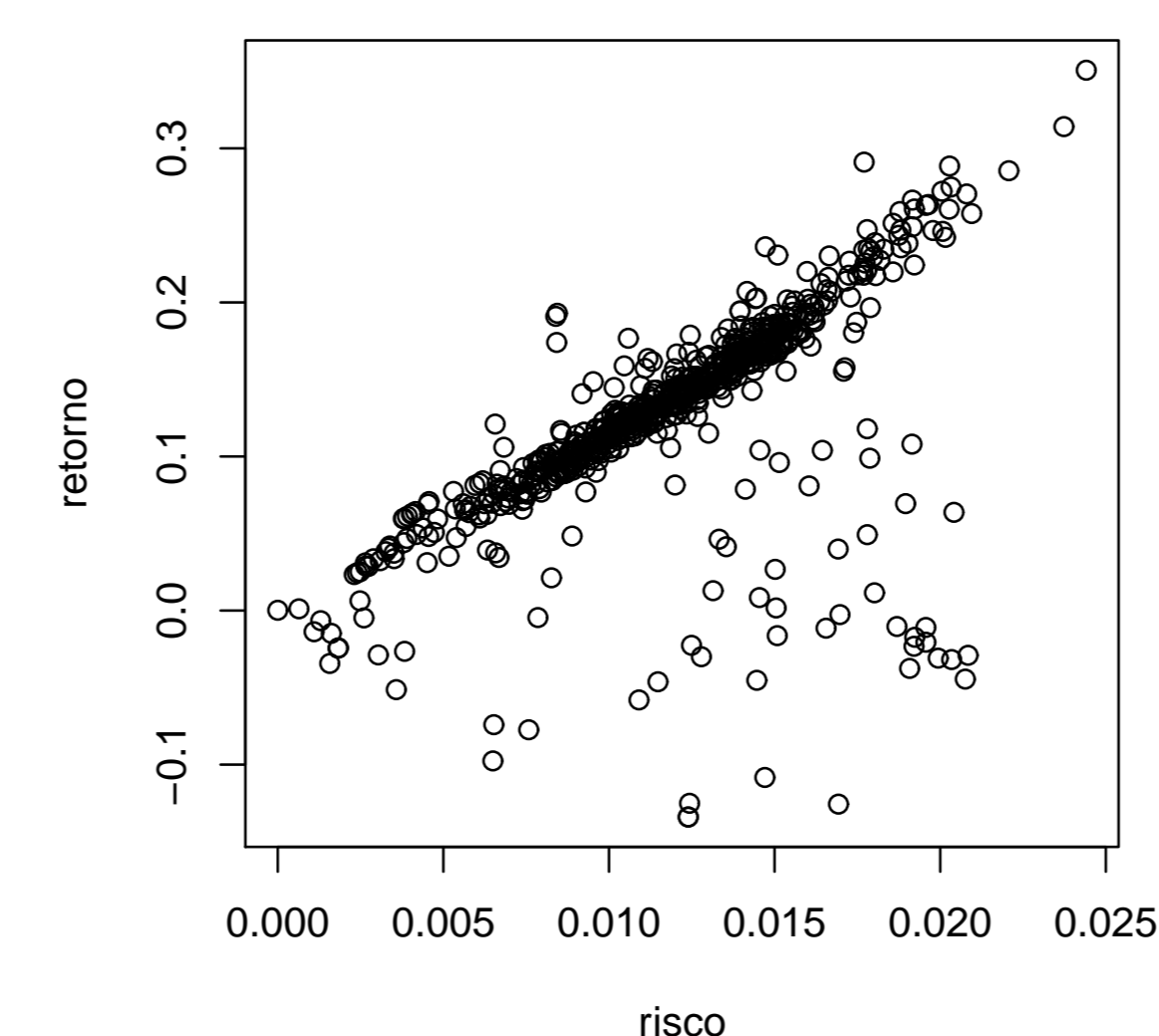


FIGURA 3: Dispersão após aplicação de Mahalanobis

Obter agrupamentos com uma tendência melhor definida pode ser extremamente útil para a determinação de um *benchmark* para a categoria, isto é, um índice que auxilie na análise comparativa de desempenho dos fundos deste grupo. Se desejado por parte do usuário final, a realocação dos *outliers* excluídos pode ser realizada através da aplicação iterativa (fundo e *benchmark* de cada categoria) de indicadores que representem a similaridade de comportamento, como o *Beta* e o *Tracking Error*, por exemplo.

## Estrutura do Projeto

O diagrama abaixo (FIGURA 4) representa uma visão geral de como o projeto foi estruturado.

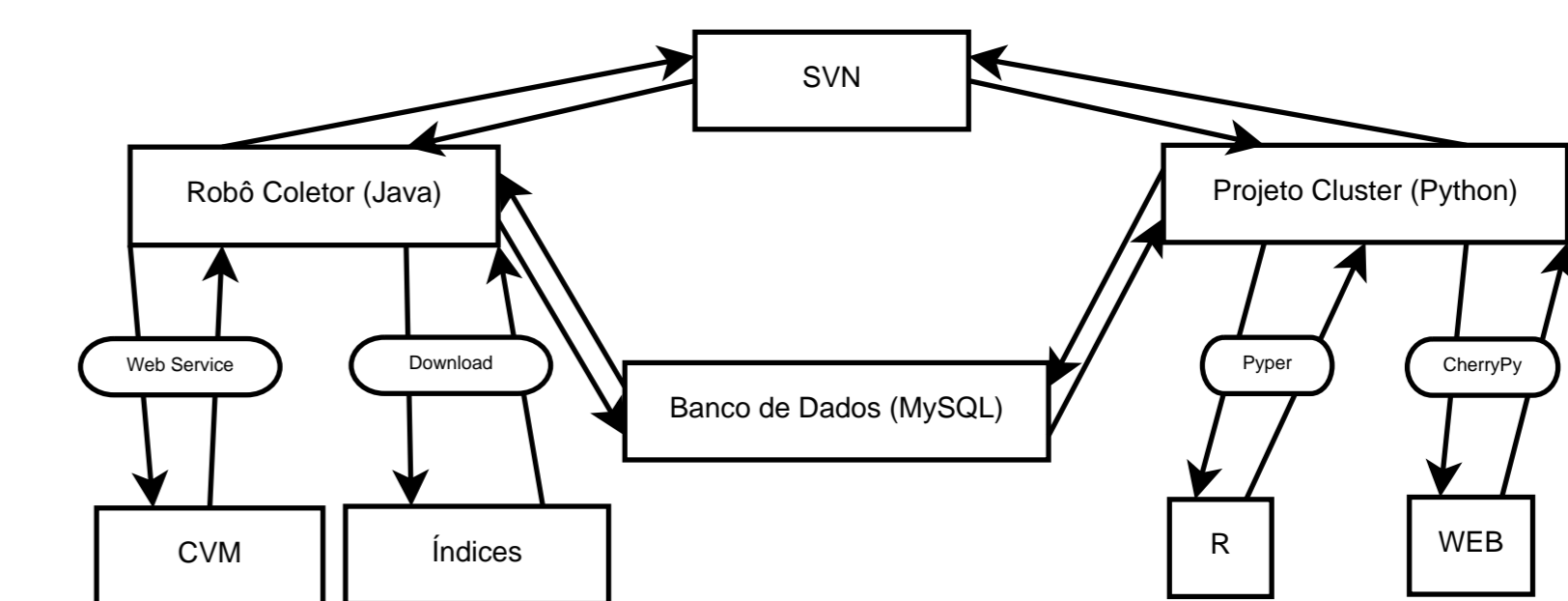


FIGURA 4: Estrutura do projeto

### Módulo - Coleta de Informações

Para a coleta do cadastro e dados históricos dos fundos, além das séries de índices, foram criados robôs na linguagem Java. No caso dos fundos, foi gerado um cliente para o consumo de um *Web Service* disponibilizado pela *CVM (Comissão de Valores Mobiliários)*. Já para os índices, foram realizados *downloads* de arquivos nas respectivas fontes, com uma posterior interpretação dos mesmos. A tecnologia utilizada para banco de dados foi MySQL.

### Módulo - Cálculo

O principal objetivo deste módulo, programado na linguagem Python, é concentrar todas as atividades de cálculo utilizadas no projeto. Além de possuir as calculadoras utilizadas direta ou indiretamente no trabalho (retorno, volatilidade, *beta*, *tracking error*, dentre outros), permite facilmente a inclusão de novos indicadores.

Outra característica é a integração direta entre as linguagens R e Python, utilizada para a execução do método *k-means*. Esta integração foi realizada através da biblioteca Pyper.

### Outros

Com o objetivo de gerar uma interface WEB simples e integrá-la ao Python, utilizamos o *framework* CherryPy. Além de não intrusivo, proporciona uma certa agilidade na programação e é extremamente leve. Já para controle de versão foi utilizado o Subversion (SVN). Além de possuir código aberto, proporciona facilidade de comunicação entre equipes de programadores por gerenciar arquivos cronologicamente.



## Referências

- [Cas10] Caio Ramos Casimiro. Desenvolvimento de uma api para estimação de betas variáveis de fundos brasileiros. 2010.
- [Eve11] Landau Everitt. *Cluster Analysis*. Wiley, 2011.
- [Had92] A. Hadi. Identifying multiple outliers in multivariate data. *Journal Royal Statistics Society*, 1992.
- [Sil07] Roberta Anchieta da Silva. Estimação dinâmica do beta do modelo CAPM em fundos de ações. 2007.