

Universidade De São Paulo  
Instituto de Matemática e Estatística  
Departamento de Ciência da Computação

MAC0499 - Trabalho de Formatura Supervisionado

# Classificação e Análise de Desempenho de Fundos

Aluno: José Corsini Filho  
Orientador: Prof. Dr. Renato Vicente

Novembro de 2012

# Fund clustering

Motivação

# Fundo de Investimento

---

- ▶ O que é?
  - ▶ Ferramenta financeira: carteira de aplicações.
- ▶ Objetivo
  - ▶ Ganhos financeiros de acordo com uma meta.
- ▶ Como?
  - ▶ Adotar estratégias de atuação para compra e venda de ativos: Ações, Títulos Públicos, Fundos, Moedas, dentre outros.
- ▶ Quem pode?
  - ▶ Depende do fundo. Muitos são abertos ao público em geral.
- ▶ Órgão Regulador
  - ▶ Comissão de Valores Mobiliários (CVM)



# Classificação

---

## ▶ Categorias

- ▶ Fundos são organizados de acordo com diversas características, dentre elas estratégia e segmento de atuação.

## ▶ Quem classifica?

- ▶ Fontes: CVM (Pública), ANBIMA (Privada), Usuário Final, dentre outros.

## ▶ Por que classificar?

- ▶ Acompanhamento/Análise do investimento.
- ▶ Verificação de enquadramento na categoria.
- ▶ Comparação ao *Benchmark*: índices (IBOVESPA, CDI, IPCA etc.), própria categoria ou outros fundos.

## ▶ Desafio

- ▶ Volume de dados (mais de 10000 fundos).
- ▶ Propor um algoritmo que minimize número de *outliers* por categoria.

# Alimentação do Banco de Dados

# Banco de Dados

---

- ▶ Tecnologia Utilizada: MySQL

- ▶ Open Source; Popular; Roda em mais de 20 plataformas.



- ▶ Forma de Coleta: Robôs

- ▶ Cada tipo de informação possui um robô diferente para busca de dados.
- ▶ Agendamento: biblioteca Quartz.

# Banco de Dados

---

- ▶ **Alimentação das Informações de Fundos:**
  - ▶ Cota, PL, Data Início, Categoria CVM, dentre outros.
  - ▶ Fonte: CVM
  - ▶ WEB SERVICE
  - ▶ Linguagem Utilizada: JAVA
  - ▶ Dificuldade: instabilidade do serviço
  
- ▶ **Alimentação das Informações de Índices**
  - ▶ IBOVESPA, PTAX, IRF-M
  - ▶ Download de Arquivos + Parser
  - ▶ Linguagem Utilizada: JAVA





Estudo  
Realizado

# Estimação Dinâmica do Beta - CAPM

---

- ▶ **Dissertação: Roberta Anchieta da Silva (2007) – IME – FEA - USP**
  - ▶ Estudo sobre o indicador Beta do modelo CAPM e a análise de sua invariância no tempo.
- ▶ **Beta ( $\beta$ )**
  - ▶ Sensibilidade de um ativo em relação ao mercado (carteira teórica): coeficiente angular da reta dada pela regressão linear de seus retornos.
  - ▶  $\beta = 1$ : ativo acompanha o mercado;
  - ▶  $\beta < 1$ : ativo oscila menos, no mesmo sentido;
  - ▶  $\beta > 1$ : ativo oscila mais, proporcionalmente ao valor obtido.
- ▶ **CAPM (Capital Asset Pricing Model)**
  - ▶ Modelo de precificação de ativos financeiros.
  - ▶ Premissa: invariância do Beta ao longo do tempo (um único Beta para todo o período).

# Estimação Dinâmica do Beta - CAPM

---

- ▶ **Modificação do modelo CAPM:**
  - ▶ Calculo de uma série de regressões (Betas).
  - ▶ Difícil conclusão.
- ▶ **Filtro de Kalman:**
  - ▶ Algoritmo recursivo de estimação.
  - ▶ Estimação de estado de um processo não observável.
  - ▶ Filtragem + Aprendizagem + Solução.
  - ▶ Conclusão: séries estacionárias, isto é, gestor tende a retornar à média.
- ▶ **Outras técnicas utilizadas:**
  - ▶ Teste de Chow: quebra estrutural (tendência estável) da série.
  - ▶ Testes ADF: raiz unitária, estudo da flutuação em cima de tendência.
  - ▶  $R^2$ : ajustamento dos retornos à regressão.

# API para estimação de Betas Variáveis

---

- ▶ **TCC: Caio Ramos Casimiro (2010) – EACH – USP**
  - ▶ Análise de Dados.
  - ▶ Estimação do Beta dinâmico dos fundos brasileiros.
  - ▶ Desenvolvido como um módulo dentro de uma empresa.
  
- ▶ **Resumo:**
  - ▶ K-means: utilizado neste projeto.
  - ▶ Análise de Componentes Principais (PCA): fatores de risco.
  - ▶ Regressões Lineares Múltiplas (uma delas a Descontada).
  - ▶ Reaproveitamento das sugestões de tecnologias utilizadas.

# Classificação (Clustering)

# Classificação - Clustering

---

- ▶ Fórmula: conhecimento técnico + criatividade
  - ▶ Combinar ferramentas de forma a obter um resultado satisfatório.
- ▶ Não determinística: depende do objetivo do usuário
  - ▶ Classificar grupos de performance.
  - ▶ Classificar grupos de risco.
  - ▶ Classificar ambos, simultaneamente.
  - ▶ Diversas áreas: *marketing*, comportamento de populações, dentre outros.
- ▶ Uma alternativa: *K-means*

# *K-means*

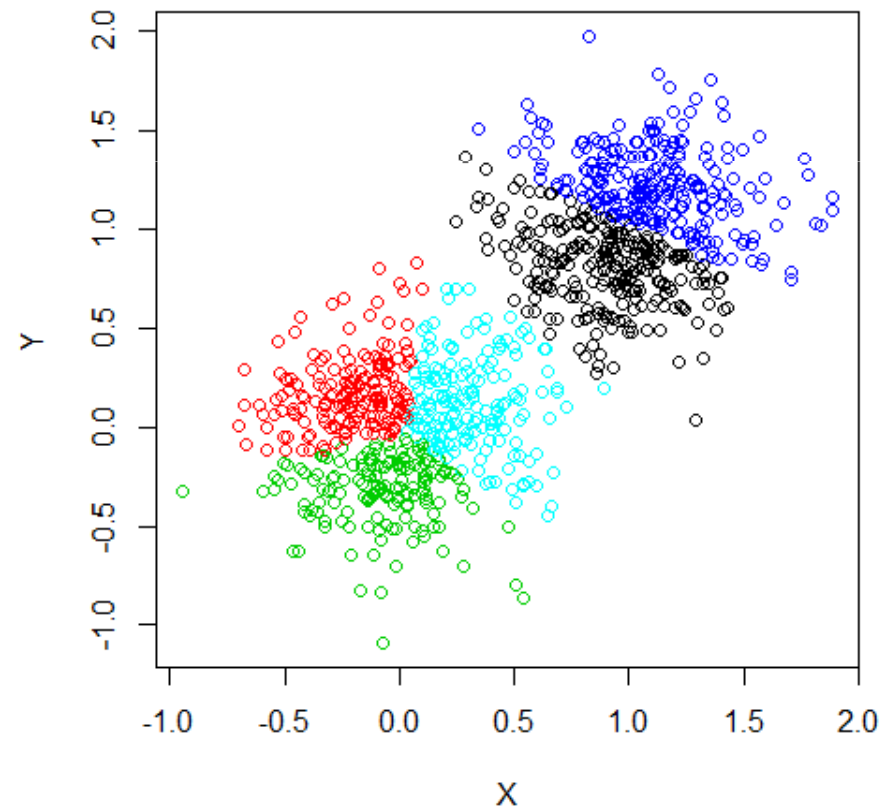
---

- ▶ J.B. MacQueen (1967)
- ▶ Técnica de agrupamento de dados
  - ▶ Utiliza os próprios dados.
  - ▶ Similaridade de grupos.
- ▶ Algoritmo
  - ▶ Definição de centróides heurísticamente (amostra).
  - ▶ Enquadramento de cada elemento ao *cluster* mais próximo (função de similaridade).
  - ▶ Os *clusters* têm os centróides recalculados e é realizada uma nova iteração do processo de enquadramento.
  - ▶ Quando parar?
    - Uma nova iteração não surte realocação de elemento;
    - Limite de iterações pré-definido.

# *K-means*: Exemplo

---

- ▶ Bi-dimensional .
- ▶ Amostra: 1000 pontos.





# Pós-Classificação

# Análise das Categorias

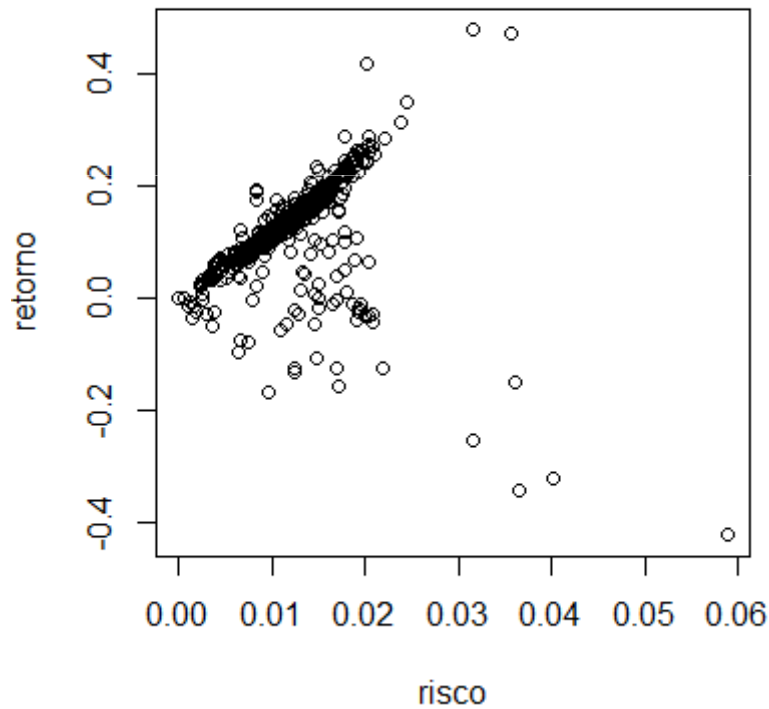
---

- ▶ Análise de dispersão dos fundos de uma categoria
  - ▶ Surgimento de *outliers* (observações numericamente distantes da tendência do grupo analisado).
- ▶ Uma solução: Distância de Mahalanobis
  - ▶ Aplicar com 2 variáveis (dispersão).
  - ▶ Baseada nas correlações entre variáveis.
  - ▶ Parametrizar de forma a remover fundos indesejados.

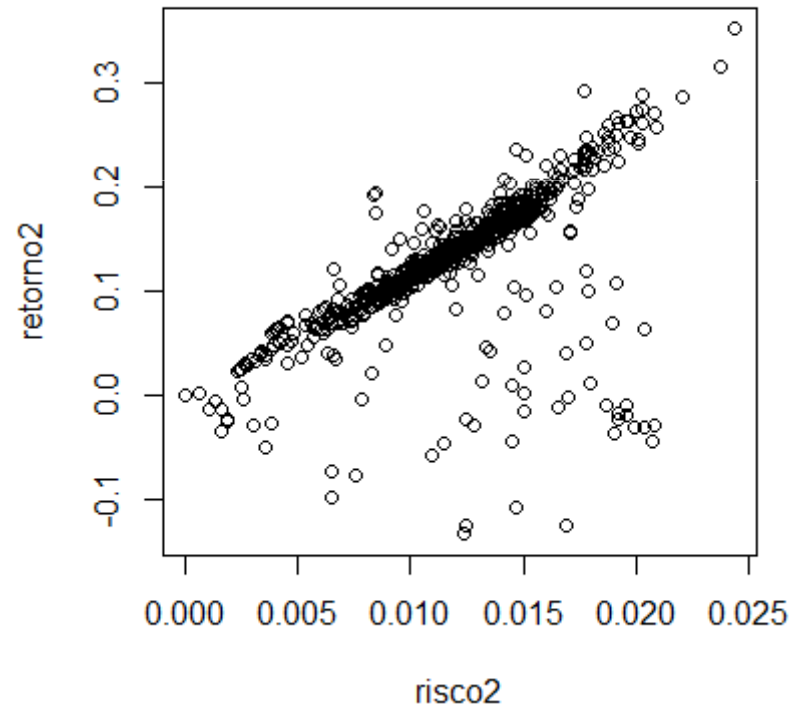
# Análise das Categorias - Exemplo

---

- ▶ Retorno x Risco (volatilidade/desvio padrão contínuo)
- ▶ Resultado Inicial:



Após aplicação de Mahalanobis:



# Análise das Categorias

---

- ▶ Atribuição de um *benchmark*
  - ▶ Tentativa de determinar uma tendência da categoria. Exemplo: Fundos de Ações, atribuir índice IBOVESPA.
  - ▶ Uma possibilidade: *Tracking Error* do retorno médio da categoria com cada um dos *benchmarks* disponíveis.
- ▶ Utilização:
  - ▶ Acompanhamento da categoria como um todo.
  - ▶ Acompanhamento individual dos fundos da categoria.
  - ▶ Tratamento de *outliers*.

# Tratamento *Outliers*

---

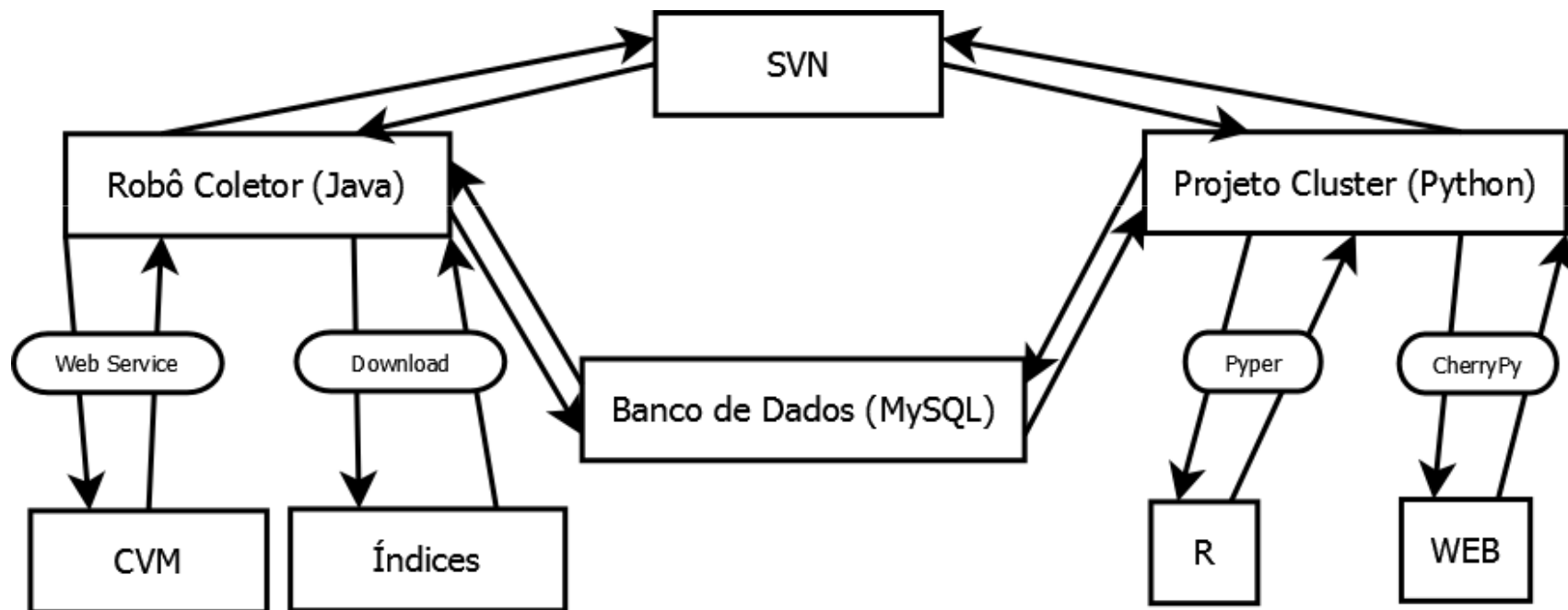
- ▶ Fundos devem ser reenquadrados
- ▶ Algumas estratégias possíveis:
  - ▶ *Tracking error*: fundo x *benchmark* das categorias.
  - ▶ *K-means* entre categorias e *outliers*?
    - Problema: um fundo poderá continuar desenquadrado.

Estrutura

# Visão Geral

---

► Mapa do projeto



# Módulos - Cálculo

---

- ▶ Estrutura de cálculo.
  - ▶ Linguagem: Python (performance).
  - ▶ Objetivo: encapsular os cálculos de indicadores realizados localmente, com base nas séries históricas (retorno, volatilidade, *tracking error*, dentre outros).
  - ▶ Característica: fácil inclusão de novos indicadores.
  
- ▶ Integração com a Linguagem R:
  - ▶ Objetivo: realizar chamadas em R, através do Python.
  - ▶ Motivação:
    - Diversos algoritmos avançados de cálculo;
    - Robustez no processamento de dados.
  - ▶ Biblioteca utilizada: Pyper.
  - ▶ Utilização: cálculo do *K-means*.





# Módulos – WEB + SVN

---

- ▶ Framework utilizado: *CherryPy*.
  - ▶ Não intrusivo.
  - ▶ Características: Facilidade + Agilidade + Leve.
  
- ▶ Controle de versão: *Subversion (SVN)*
  - ▶ Código aberto.
  - ▶ Gerenciamento de arquivos: cronológico.



SUBVERSION®

# Fim

---

- ▶ Contato: [prof.corsini@gmail.com](mailto:prof.corsini@gmail.com)
- ▶ Página Trabalho: <http://www.linux.ime.usp.br/~ze/mac499/>