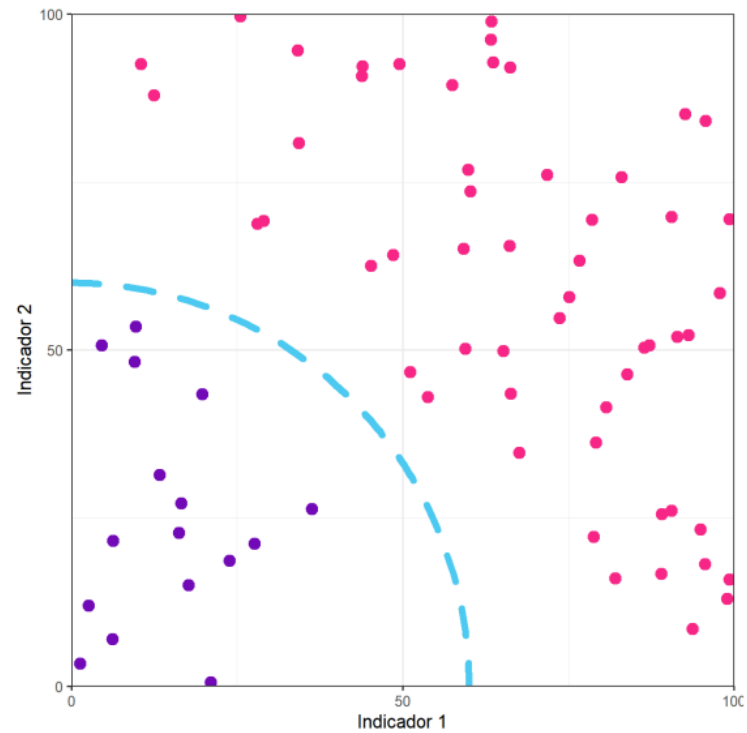


# Support Vector Machines

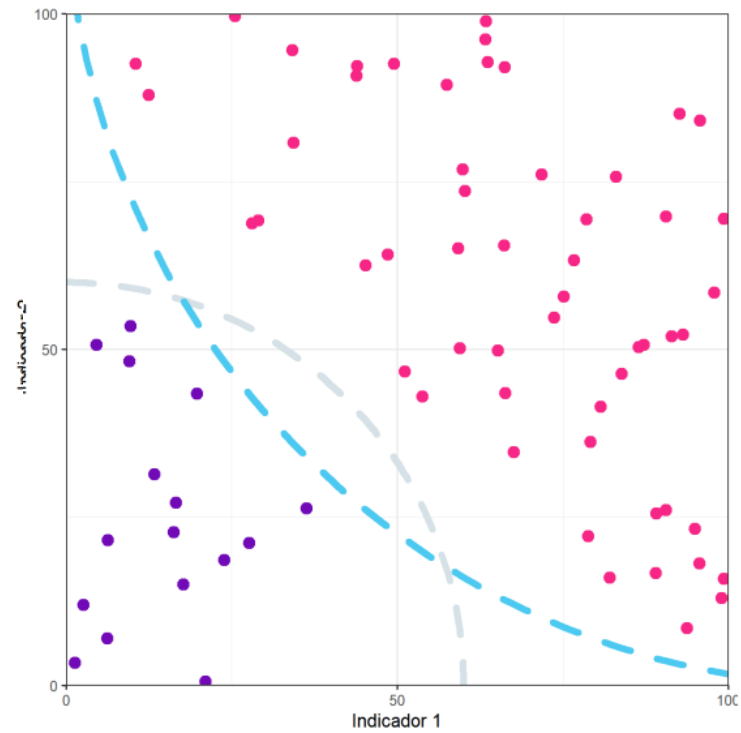


INFORMAÇÃO,  
TECNOLOGIA  
& INOVAÇÃO

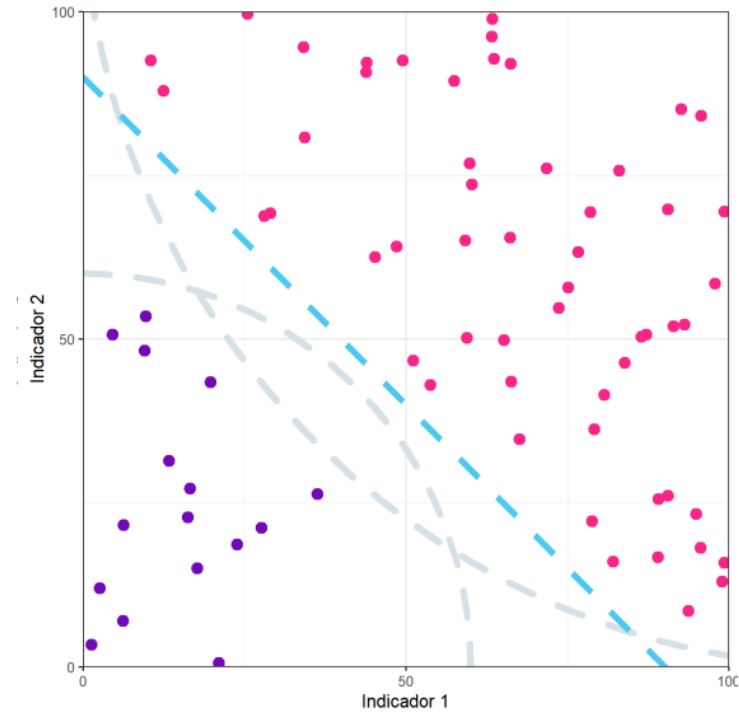
# SVM



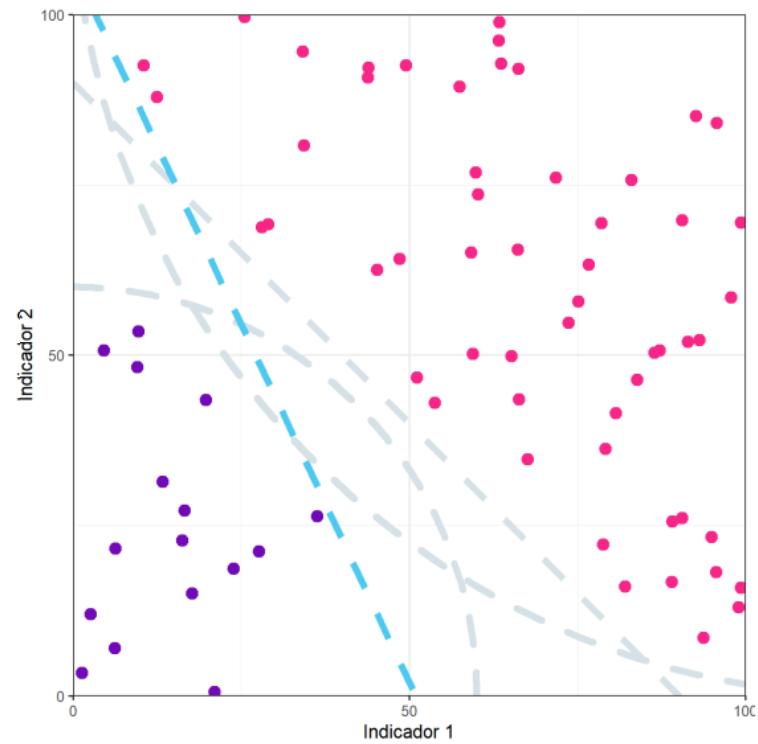
# SVM



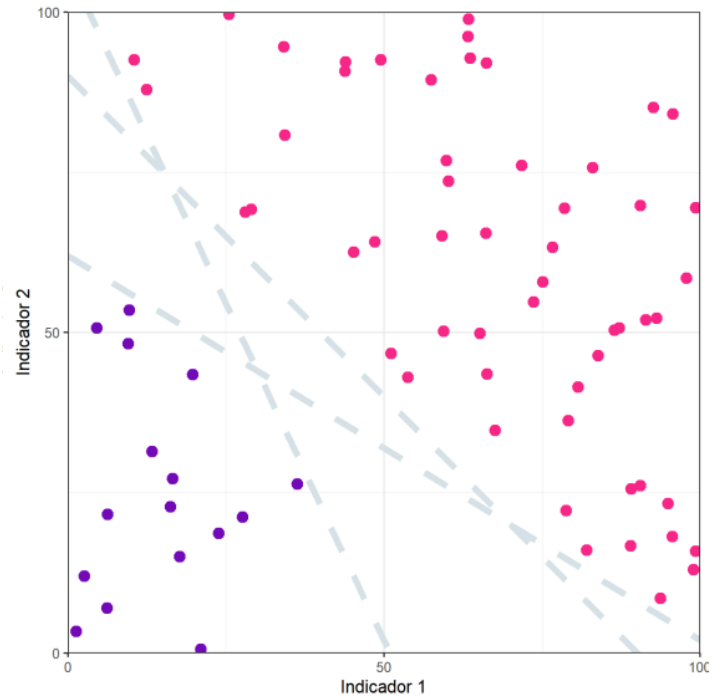
# SVM



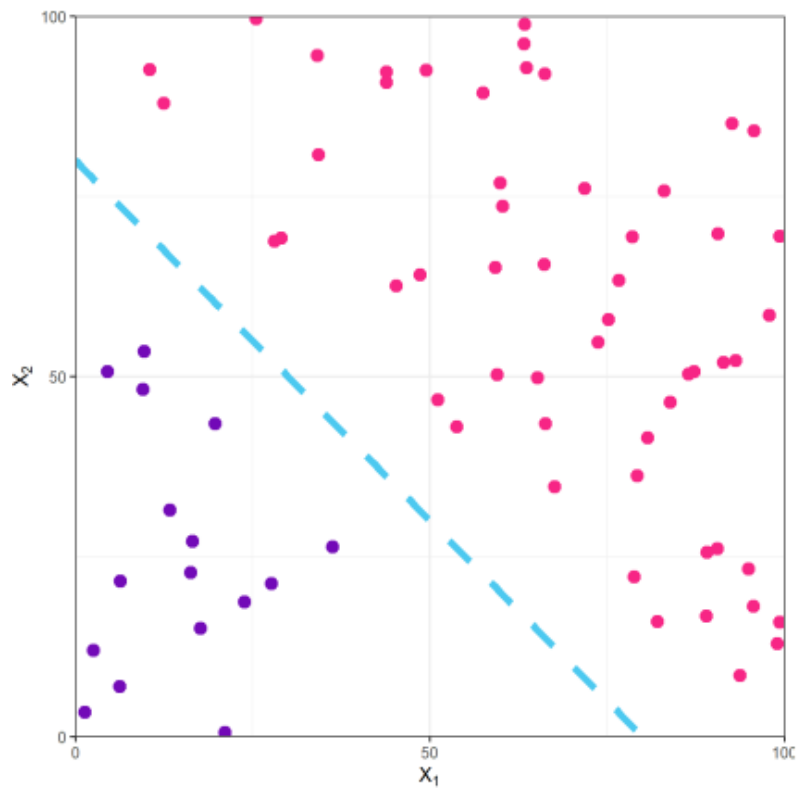
# SVM



# Retas



# Melhor reta



# Melhor reta

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} > 0 \text{ se } y_i = 1$$

e

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} < 0 \text{ se } y_i = -1.$$

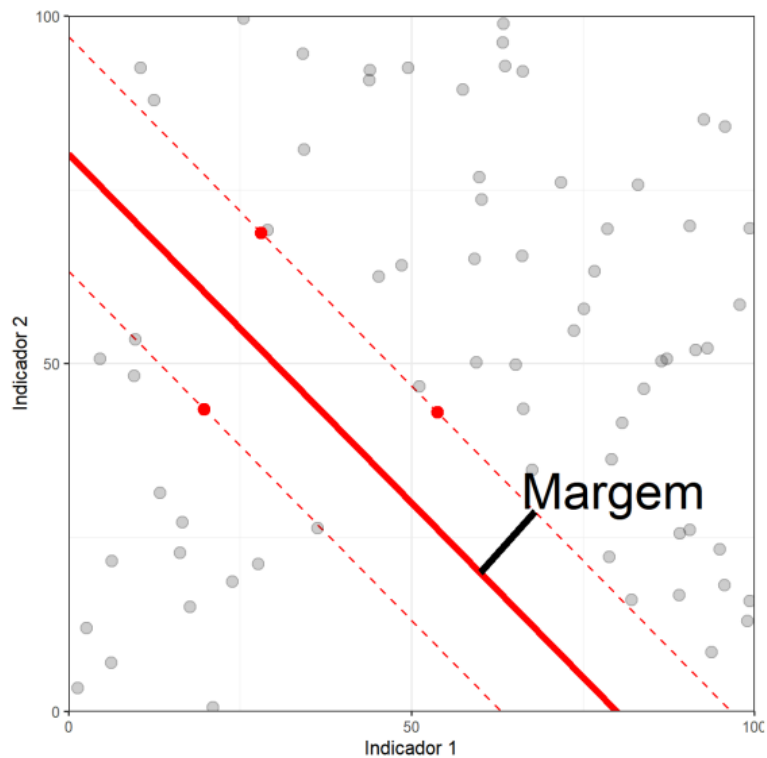
Note que as duas equações acima podem ser resumidas em

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) > 0 \text{ para } i = 1, \dots, n.$$





# Melhor reta



# Melhor reta

O hiperplano de margem máxima é dado pela solução do seguinte objetivo:

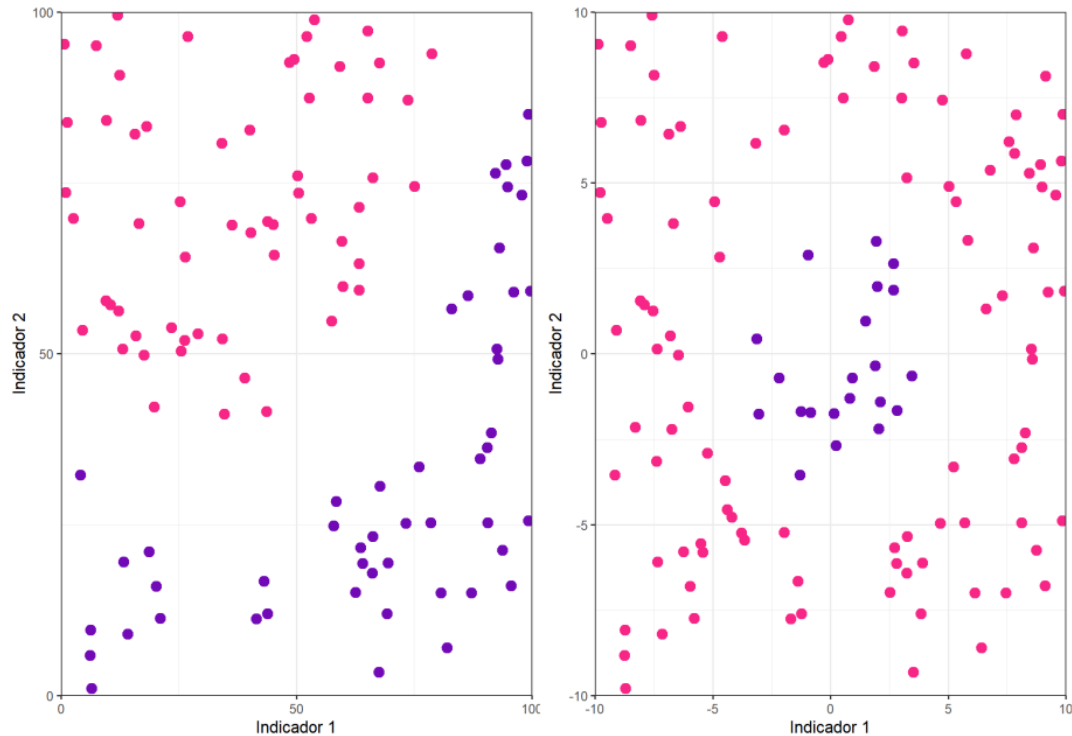
$$\underset{\beta_0, \beta_1, \dots, \beta_p}{\text{maximizar}} \ M$$

sujeito a

$$\sum_{j=1}^p \beta_j^2 = 1 \text{ e } y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M \text{ para } i = 1, \dots, n.$$



# Caso não separável



# Caso não separável

$$\underset{\beta_0, \beta_1, \dots, \beta_p, \epsilon_1, \dots, \epsilon_n}{\text{maximizar}} \quad M$$

sujeito a

$$\sum_{j=1}^p \beta_j^2 = 1 \text{ e } y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}) \geq M(1 - \epsilon_i)$$

com  $\epsilon_i \geq 0$  e  $\sum_{i=1}^n \epsilon_i \leq C$ .

Aqui  $C$  é um hiperparâmetro e  $M$  é a margem.

As variáveis  $\epsilon_i$ 's são variáveis **slack** que permitem que as observações individualmente estejam no lado errado da margem ou do hiperplano.

- $\epsilon = 0$  : observação do lado correto da margem.
- $\epsilon > 0$  : observação do lado errado da margem.
- $\epsilon > 1$  : observação do lado errado do hiperplano.



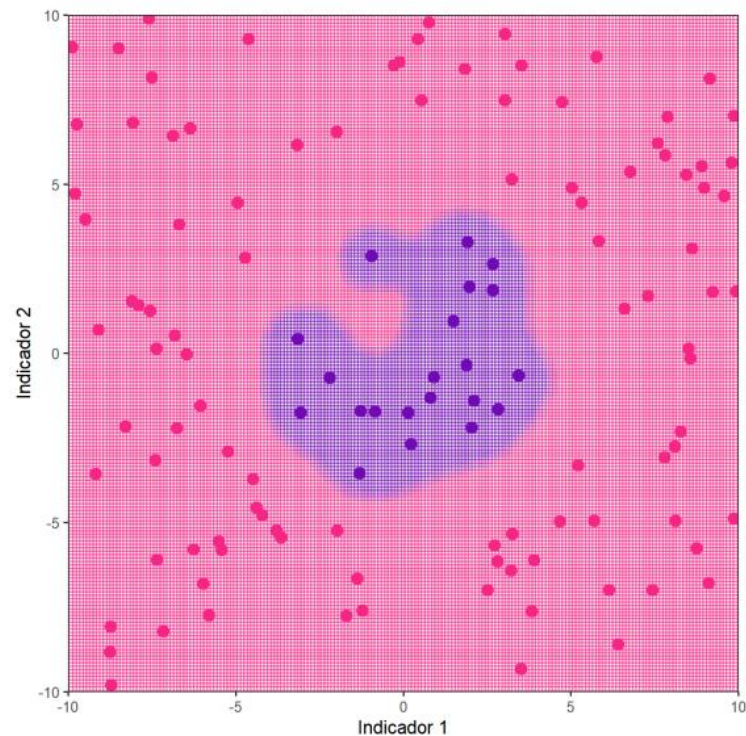
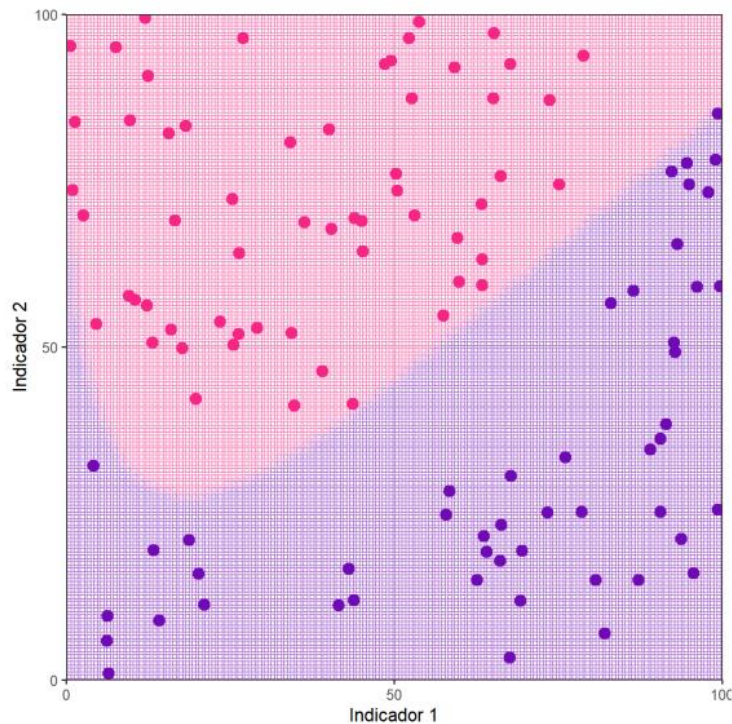
# Caso não separável

É interessante notar que  $C$  define o *trade-off* viés variância.

- $C$  pequeno: margens estreitas que raramente são violadas. São altamente dependentes dos dados. Apresentam **viés pequeno** e **variância alta**.
- $C$  grande: margens largas, menos dependente dos dados. Apresentam **viés alto** e **variância baixa**.



# Fronteira Não-linear



# Kernels

- **Gaussian RBF kernel:**  $k(x, x') = \exp(-\sigma ||x - x'||^2)$
- **Polynomial kernel:**  $k(x, x') = (\text{scale} \langle x, x' \rangle + \text{offset})^{\text{degree}}$
- **Linear kernel:**  $k(x, x') = \langle x, x' \rangle$
- **Hyperbolic tangent kernel:**  $k(x, x') = \tanh(\text{scale} \langle x, x' \rangle + \text{offset})$
- **Laplacian kernel:**  $k(x, x') = \exp(-\sigma ||x - x'||)$
- **Bessel kernel:**  $k(x, x') = (-\text{Bessel}_{(\nu+1)}^n \sigma ||x - x'||^2)$
- **ANOVA RBF kernel:**  $k(x, x') = \sum_{1 \leq i_1 \dots \leq i_D \leq N} \prod_{d=1}^D k(x_{id}, x'_{id})$  em que  $K(x, x')$  é um kernel *Gaussian RBF*.



# Amazon





## Tarefa - IML 1.2



# Abandono de clientes (churn)

A análise de abandono de clientes (churn) é um importante instrumento para a gestão estratégica e operacional de um negócio. O objetivo prever se um cliente irá ou não abandonar um serviço e entender quais fatores impactam para esta decisão. Nesta tarefa, iremos elaborar um modelo de predição para churn de clientes de uma empresa provedora de internet.

Características do banco de dados:

- 72274 observações
- 10 colunas (9 features e 1 variável resposta)
- Dados de características dos serviços contratados pelos clientes
- Indicadora de abandono

Na sua tarefa você vai ter que criar um modelo de classificação para prever se haverá ou não abandono de um cliente de acordo com as suas características.

Você tem total liberdade para selecionar os modelos que deseja testar, criar novas covariáveis, remover covariáveis e selecionar técnicas de comparação entre os modelos que você escolher

# Pontos importantes!

## Análise descritiva

- A análise descritiva faz parte do processo de modelagem
- Entenda os dados
- Identifiquem o que é variável resposta e o que é variável explicativa
- Identifique missing, outliers e outros problemas de dados

## Modelagem

- Entendam qual técnica de aprendizado faz mais sentido para o dados
- Testem diferentes modelos
- Avaliem os modelos por alguma medida de comparação
- Seleccionem variáveis de forma inteligente
- Conclua qual modelo fornece o melhor ajuste

## Estrutura

- Utilize a metodologia CRISP-DM
- Motive o desenvolvimento da análise (*business case*)
- Siga uma ordem lógica no desenvolvimento
- Faça comentários explicando seu raciocínio (o porquê de utilizar determinada função, alguma característica importante dos dados, motivação para próximos passos)
- Explícite a interpretação do modelo
- Retire comentários desnecessários ou que não impactem diretamente o código ou a interpretação
- Façam a atividade utilizando o jupyter notebook