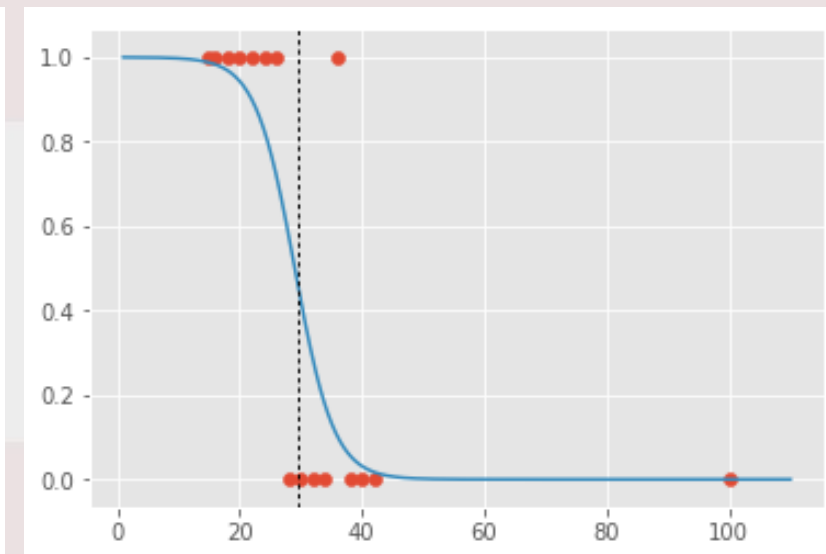
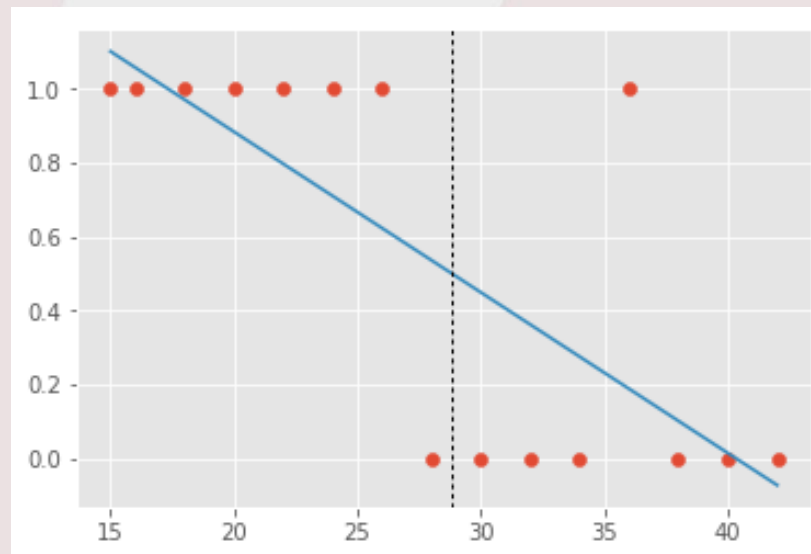
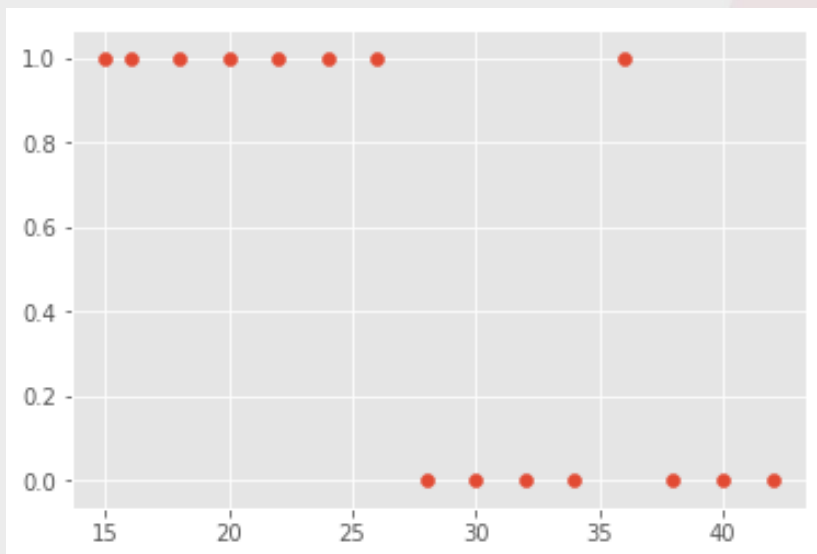


Regressão Logística

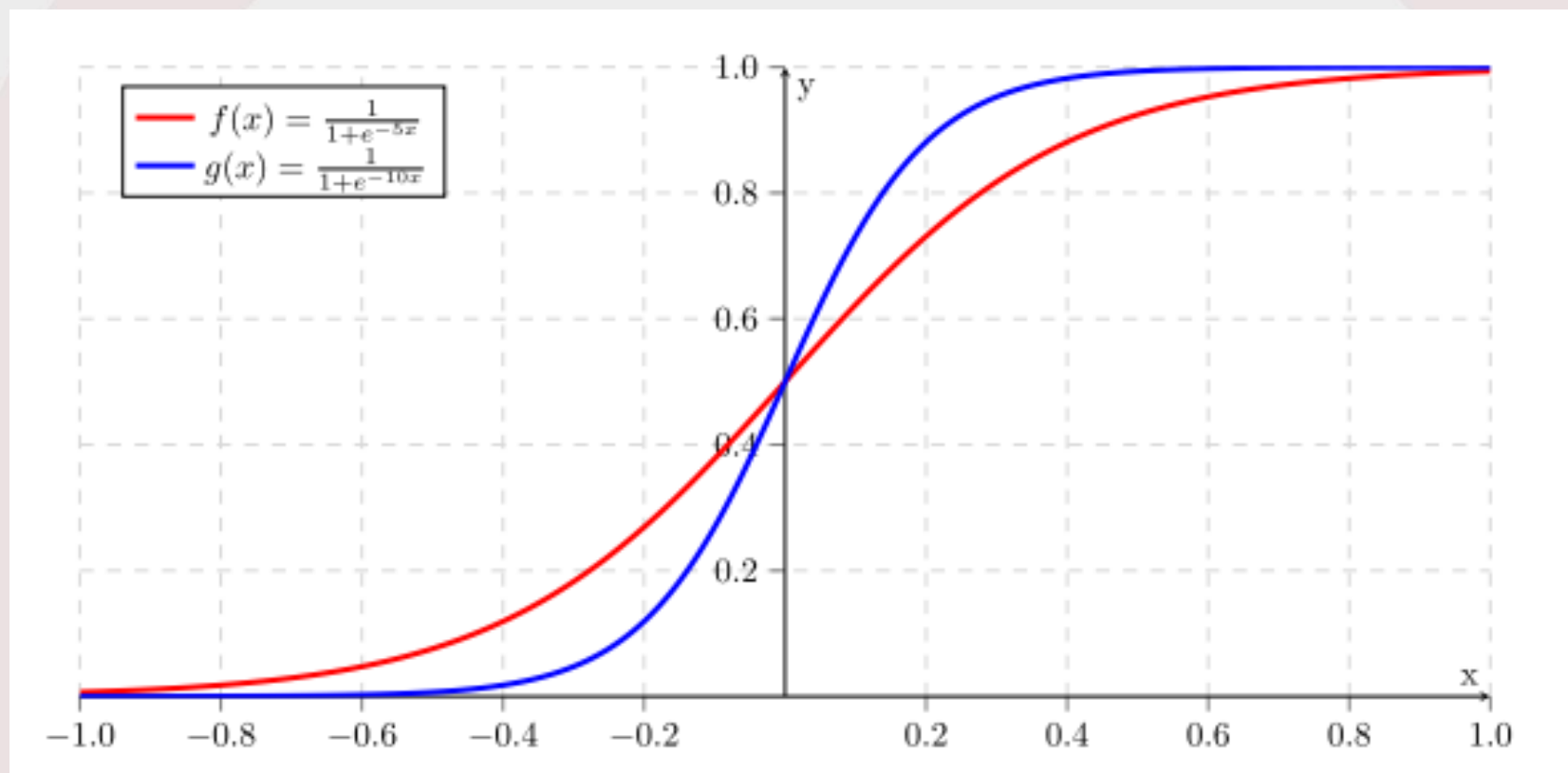
- A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias.



Regressão Logística

- Função Logística:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



Regressão Logística

- Variável Y tem distribuição binomial e binária.
- Usa o método da máxima verossimilhança.
- Precisa apenas de um pequeno número de suposições para realizar as previsões.
- Retorna a probabilidade.

Regressão Logística

- **Regressão logística binominal**

- No modelo de regressão logística binominal, **os objetos são classificados em dois grupos ou categorias.**

- **Regressão logística ordinal**

- O modelo de regressão logística ordinal é diferente porque **trabalha com o conceito de categorias ordenadas.**
- Neste caso, os objetos são classificados em três ou mais classes que possuem uma ordem já determinada.

- **Regressão logística multinomial**

- No modelo de regressão logística multinomial, os objetos são classificados em **três ou mais categorias que não possuem ordem** entre si.

Regressão Logística

- Prova: CESPE / CEBRASPE - 2016 - FUNPRES-EXE - Analista - Área Investimentos
- Um modelo de regressão logística só aceita variáveis categóricas; um modelo de regressão linear só aceita variáveis quantitativas.

Regressão Logística

- Prova: CESPE / CEBRASPE - 2016 - FUNPRES-P-EXE - Analista - Área Investimentos
- Um modelo de regressão logística só aceita variáveis categóricas; um modelo de regressão linear só aceita variáveis quantitativas.

Regressão Logística

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Uma determinada repartição pública fez um levantamento do tempo y , em minutos, que os cinco funcionários de uma sessão gastam para chegar ao trabalho em função da distância x , em quilômetros, de suas residências. O resultado da pesquisa realizada com cada um deles é apresentado na tabela a seguir, em que respectivamente, as médias amostrais das variáveis x e y .
- Uma forma de melhorar o modelo de regressão linear para a situação em questão é utilizar o modelo de regressão logística, uma vez que a variável dependente se apresenta de forma quantitativa.

i	tempo y_i	distância x_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \times (y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	10	5	-4	-7	28	16
2	20	5	-4	3	-12	16
3	15	10	1	-2	-2	1
4	10	10	1	-7	-7	1
5	30	15	6	13	78	36
média	17	9				

Regressão Logística

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- Uma determinada repartição pública fez um levantamento do tempo y , em minutos, que os cinco funcionários de uma sessão gastam para chegar ao trabalho em função da distância x , em quilômetros, de suas residências. O resultado da pesquisa realizada com cada um deles é apresentado na tabela a seguir, em que respectivamente, as médias amostrais das variáveis x e y .
- Uma forma de melhorar o modelo de regressão linear para a situação em questão é utilizar o modelo de regressão logística, uma vez que a variável dependente se apresenta de forma quantitativa.

i	tempo y_i	distância x_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \times (y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	10	5	-4	-7	28	16
2	20	5	-4	3	-12	16
3	15	10	1	-2	-2	1
4	10	10	1	-7	-7	1
5	30	15	6	13	78	36
média	17	9				

Regressão Logística

- Prova: FGV - 2010 - FIOCRUZ - Tecnologista em Saúde - Estatística
- Assinale a alternativa que indique o problema mais apropriado para aplicação da regressão logística.
- A Para obter o risco relativo de se desenvolver a diabetes tipo 2, em um período de 10 anos, associado com o peso do indivíduo e outros fatores de risco.
- B Para descrever o tamanho esperado de crianças com menos de um ano, de acordo com sua idade em meses.
- C Para prever o tempo de sobrevivência de pacientes de câncer de pulmão, de acordo com características clínicas do paciente.
- D Para descrever a distribuição de pesos de indivíduos do sexo feminino em uma certa comunidade
- E Para prever o número de casos de uma doença em diferentes municípios de acordo com algumas variáveis populacionais e epidemiológicas.

Regressão Logística

- Prova: FGV - 2010 - FIOCRUZ - Tecnologista em Saúde - Estatística
- Assinale a alternativa que indique o problema mais apropriado para aplicação da regressão logística.
- A Para obter o risco relativo de se desenvolver a diabetes tipo 2, em um período de 10 anos, associado com o peso do indivíduo e outros fatores de risco.
- B Para descrever o tamanho esperado de crianças com menos de um ano, de acordo com sua idade em meses.
- C Para prever o tempo de sobrevivência de pacientes de câncer de pulmão, de acordo com características clínicas do paciente.
- D Para descrever a distribuição de pesos de indivíduos do sexo feminino em uma certa comunidade
- E Para prever o número de casos de uma doença em diferentes municípios de acordo com algumas variáveis populacionais e epidemiológicas.

Regressão Logística

- Prova: CESPE / CEBRASPE - 2021 - SEFAZ-AL - Auditor Fiscal de Finanças e Controle de Arrecadação da Fazenda Estadual
- A regressão logística é um modelo de regressão no qual a relação entre as variáveis independentes e a variável dependente é representada por uma função degrau, a qual, por sua vez, pode ser representada por uma spline.

Regressão Logística

- Prova: CESPE / CEBRASPE - 2021 - SEFAZ-AL - Auditor Fiscal de Finanças e Controle de Arrecadação da Fazenda Estadual
- A regressão logística é um modelo de regressão no qual a relação entre as variáveis independentes e a variável dependente é representada por uma função degrau, a qual, por sua vez, pode ser representada por uma spline.

Regressão Logística

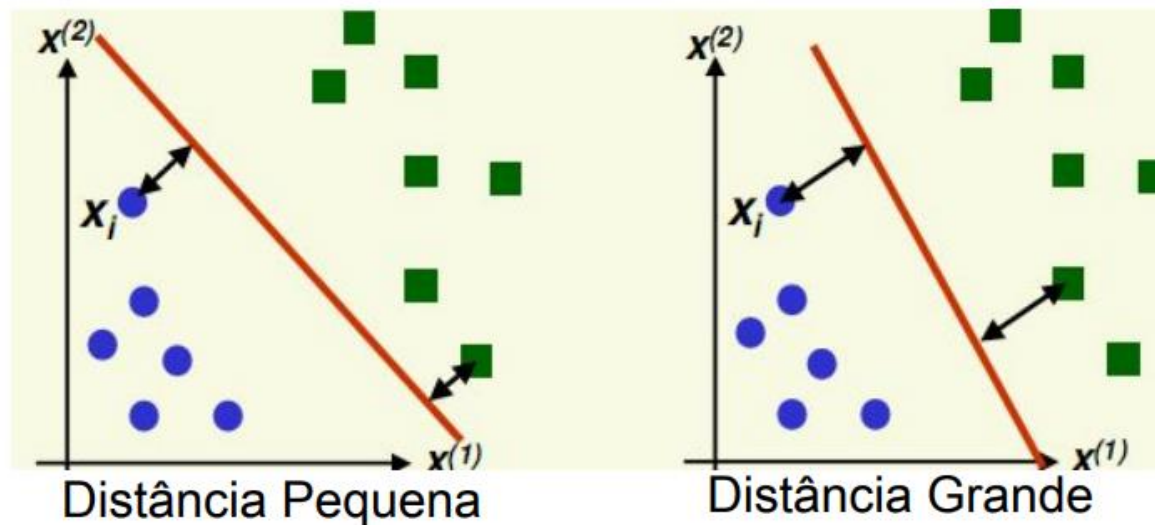
- Prova: CESPE / CEBRASPE - 2016 - FUNPRESP-EXE - Analista - Área Investimentos
- Os parâmetros de um modelo de regressão linear e logística podem ser estimados por meio da técnica de máxima verossimilhança.

Regressão Logística

- Prova: CESPE / CEBRASPE - 2016 - FUNPRESP-EXE - Analista - Área Investimentos
- Os parâmetros de um modelo de regressão linear e logística podem ser estimados por meio da técnica de máxima verossimilhança.

SVM

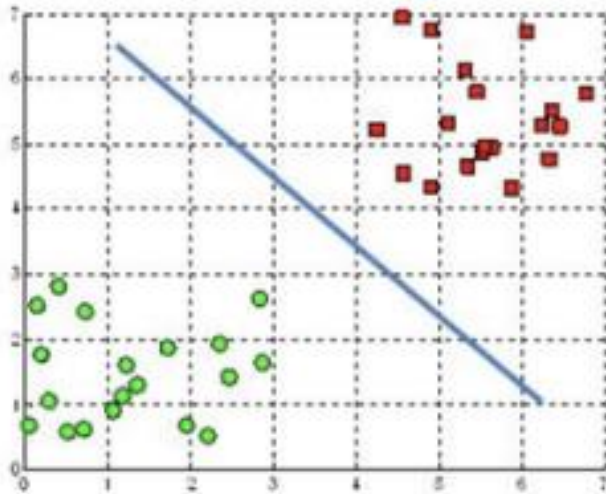
- O conceito por trás do SVM é a maximização da margem, ou seja, maximizar a distância da margem dos dados de treinamento:



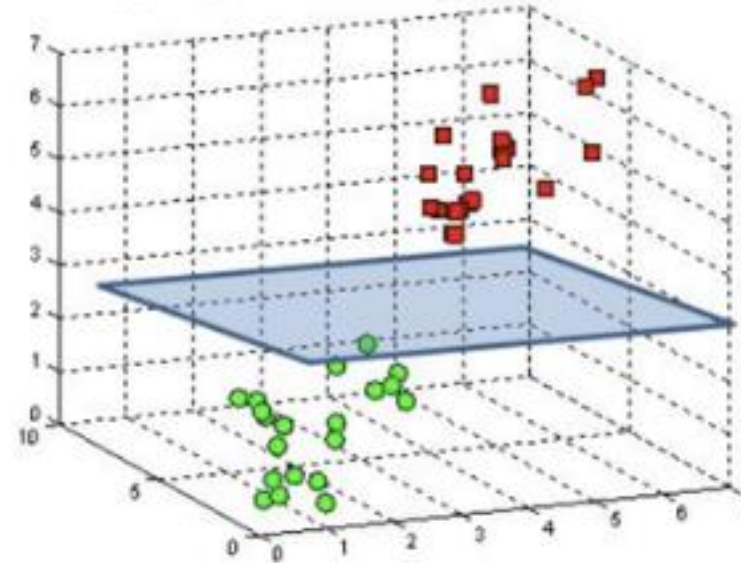
Hiperplano ótimo: Distância da margem para o exemplo da classe positiva é igual a distância da margem para o exemplo da classe negativa.

SVM

A hyperplane in \mathbb{R}^2 is a line

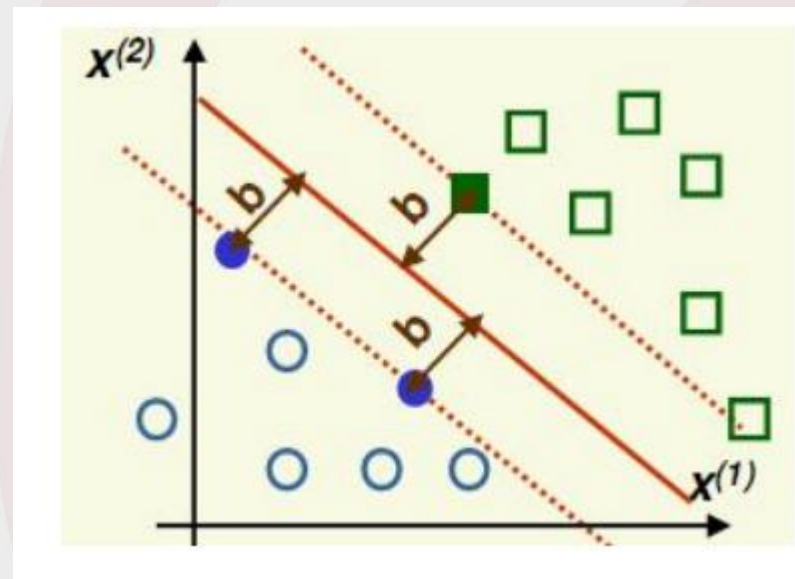


A hyperplane in \mathbb{R}^3 is a plane



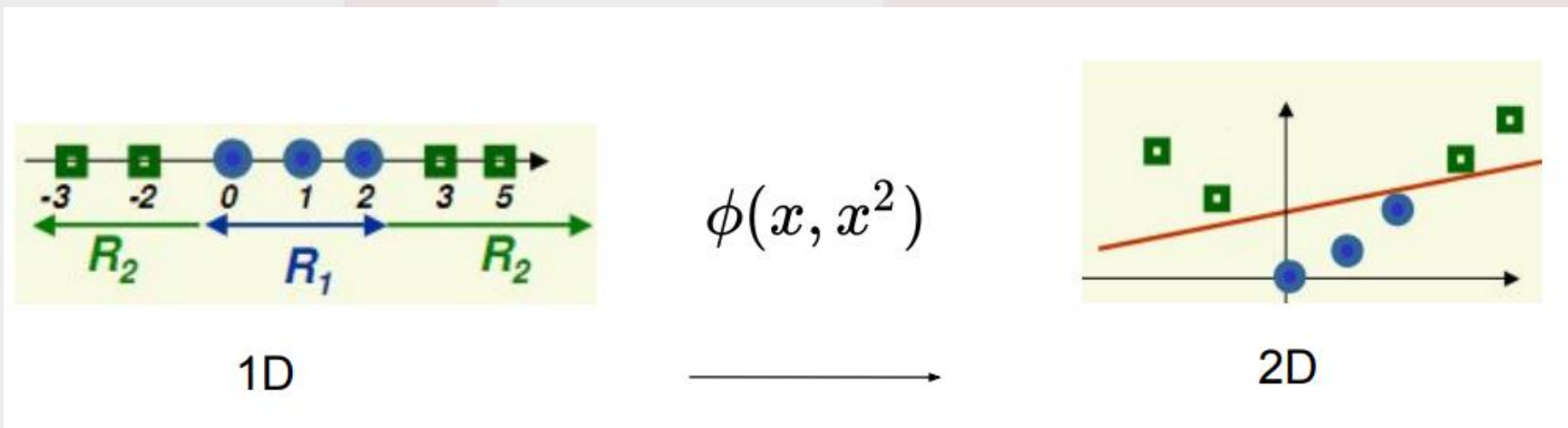
SVM

- Vetores de suporte:
 - São os exemplos da base de treinamento mais próximos do hiperplano.
 - O hiperplano é definido unicamente pelos vetores de suporte, os quais são encontrados durante o treinamento.



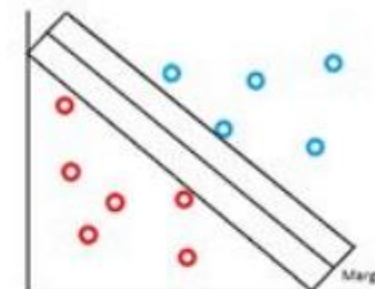
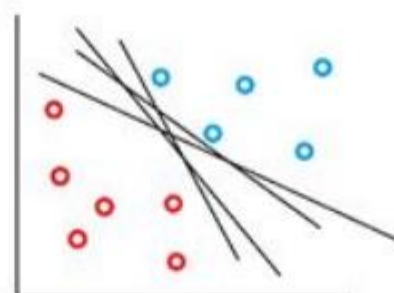
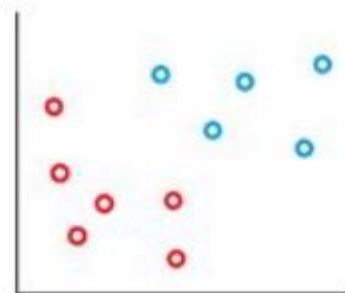
SVM

- A grande maioria dos problemas reais não são linearmente separáveis.
- A pergunta então é: o “Como resolver problemas que não são linearmente separáveis com um classificador linear?”
 - Projetar os dados em um espaço onde os dados são linearmente separáveis.
 - Projetar os dados em outra dimensão usando uma função de kernel (kernel trick).
 - Encontrar um hiperplano que separe os dados nesse espaço.

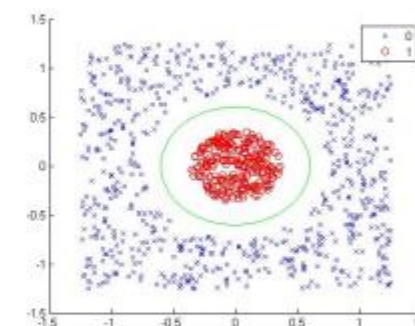
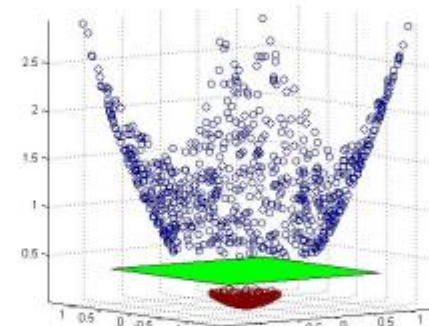
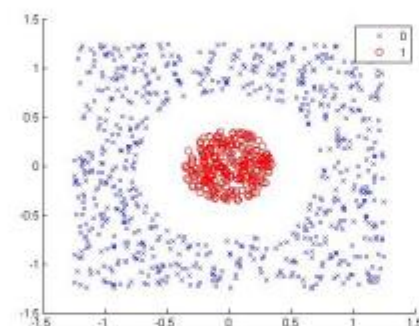


SVM

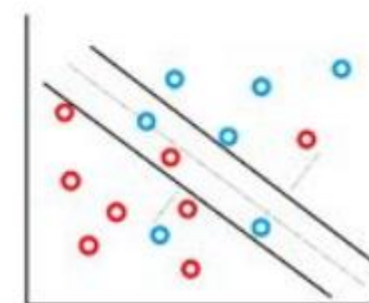
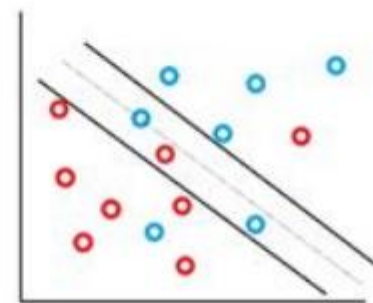
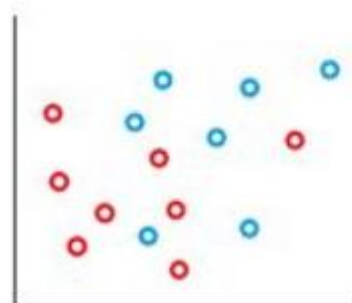
Linear (1963)



Kernel Trick (1992)



Soft Margin (1995)



SVM

- Prova: Instituto Consulplan - 2021 - SEFAZ-PI - Analista de Banco de Dados
- Considerando os métodos de Mineração de Dados, analise a seguinte descrição: “constrói os denominados classificadores lineares, que separam o conjunto de dados por meio de um hiperplano, sendo considerado um dos mais efetivos para a tarefa de classificação.” Trata-se de:
 - A Wang-Mendel.
 - B Backpropagation.
 - C SVM (Support Vector Machines).
 - D Classificador Bayesiano Ingênuo.

SVM

- Prova: Instituto Consulplan - 2021 - SEFAZ-PI - Analista de Banco de Dados
- Considerando os métodos de Mineração de Dados, analise a seguinte descrição: “constrói os denominados classificadores lineares, que separam o conjunto de dados por meio de um hiperplano, sendo considerado um dos mais efetivos para a tarefa de classificação.” Trata-se de:
 - A Wang-Mendel.
 - B Backpropagation.
 - C SVM (Support Vector Machines).
 - D Classificador Bayesiano Ingênuo.

SVM

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- As máquinas de vetores de suporte (SVMs) são originalmente utilizadas para a classificação de dados em duas classes, ou seja, na geração de dicotomias. Nas SVMs com margens rígidas, conjuntos de treinamento linearmente separáveis podem ser classificados. Acerca das características das SVMs com margens rígidas, julgue o item a seguir.
- Um conjunto linearmente separável é composto por exemplos que podem ser separados por pelo menos um hiperplano. As SVMs lineares buscam o hiperplano ótimo segundo a teoria do aprendizado estatístico, definido como aquele em que a margem de separação entre as classes presentes nos dados é minimizada.

SVM

- Prova: CESPE / CEBRASPE - 2022 - Petrobras - Ciência de Dados
- As máquinas de vetores de suporte (SVMs) são originalmente utilizadas para a classificação de dados em duas classes, ou seja, na geração de dicotomias. Nas SVMs com margens rígidas, conjuntos de treinamento linearmente separáveis podem ser classificados. Acerca das características das SVMs com margens rígidas, julgue o item a seguir.
- Um conjunto linearmente separável é composto por exemplos que podem ser separados por pelo menos um hiperplano. As SVMs lineares buscam o hiperplano ótimo segundo a teoria do aprendizado estatístico, definido como aquele em que a margem de separação entre as classes presentes nos dados é minimizada.

SVM

- Prova: CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados
- Em razão de seu bom desempenho, o algoritmo SVM (support vector machines) é invariante à escala dimensional dos conjuntos de dados, o que torna dispensável a padronização e o pré-processamento dos dados.

SVM

- Prova: CESPE / CEBRASPE - 2021 - SERPRO - Analista - Especialização: Ciência de Dados
- Em razão de seu bom desempenho, o algoritmo SVM (support vector machines) é invariante à escala dimensional dos conjuntos de dados, o que torna dispensável a padronização e o pré-processamento dos dados.

SVM

- Prova: COMPERVE - 2017 - UFRN - Engenheiro - Neuroengenharia
- Considere a existência de uma base de dados de sinais colhidos em experimentos de ressonância magnética funcional. Esses dados são pareados com uma variável comportamental como, por exemplo, se o sujeito estivesse em estado de vigília ou em sono profundo. Deseja-se construir um classificador capaz de identificar o estado comportamental a partir de amostras experimentais. A ferramenta de mineração de dados mais adequada para implementar esse classificador é:
- A Máquina de Vetor de Suporte, por ser uma ferramenta de aprendizado não-supervisionado.
- B Máquina de Vetor de Suporte, por ser uma ferramenta de aprendizado supervisionado.
- C Análise de Componentes Principais, por ser uma ferramenta de aprendizado supervisionado.
- D Análise de Componentes Principais, por ser uma ferramenta de aprendizado não-supervisionado.

SVM

- Prova: COMPERVE - 2017 - UFRN - Engenheiro - Neuroengenharia
- Considere a existência de uma base de dados de sinais colhidos em experimentos de ressonância magnética funcional. Esses dados são pareados com uma variável comportamental como, por exemplo, se o sujeito estivesse em estado de vigília ou em sono profundo. Deseja-se construir um classificador capaz de identificar o estado comportamental a partir de amostras experimentais. A ferramenta de mineração de dados mais adequada para implementar esse classificador é:
- A Máquina de Vetor de Suporte, por ser uma ferramenta de aprendizado não-supervisionado.
- B Máquina de Vetor de Suporte, por ser uma ferramenta de aprendizado supervisionado.
- C Análise de Componentes Principais, por ser uma ferramenta de aprendizado supervisionado.
- D Análise de Componentes Principais, por ser uma ferramenta de aprendizado não-supervisionado.

SVM

- Prova: CESGRANRIO - 2018 - Petrobras - Analista de Sistemas Júnior - Processos de Negócio
- Dois funcionários de uma empresa de crédito discutiam sobre quais algoritmos deveriam usar para ajudar a classificar seus clientes como bons ou maus pagadores. A empresa possui, para todos os empréstimos feitos no passado, um registro formado pelo conjunto de informações pessoais sobre o cliente e de como era composta a dívida inicial. Todos esses registros tinham classificações de bons ou maus pagadores, de acordo com o perfil de pagamento dos clientes. A partir desses dados, os funcionários querem construir um modelo, por meio de aprendizado de máquina, que classifique os novos clientes, que serão descritos por registros com o mesmo formato.
- A melhor opção, nesse caso, é usar um algoritmo
- A supervisionado, como SVM
- B supervisionado, como K-means
- C não supervisionado, como regressão linear
- D não supervisionado, como árvores de decisão
- E semi-supervisionado, como redes bayesianas

SVM

- Prova: CESGRANRIO - 2018 - Petrobras - Analista de Sistemas Júnior - Processos de Negócio
- Dois funcionários de uma empresa de crédito discutiam sobre quais algoritmos deveriam usar para ajudar a classificar seus clientes como bons ou maus pagadores. A empresa possui, para todos os empréstimos feitos no passado, um registro formado pelo conjunto de informações pessoais sobre o cliente e de como era composta a dívida inicial. Todos esses registros tinham classificações de bons ou maus pagadores, de acordo com o perfil de pagamento dos clientes. A partir desses dados, os funcionários querem construir um modelo, por meio de aprendizado de máquina, que classifique os novos clientes, que serão descritos por registros com o mesmo formato.
- A melhor opção, nesse caso, é usar um algoritmo
 - A supervisionado, como SVM
 - B supervisionado, como K-means
 - C não supervisionado, como regressão linear
 - D não supervisionado, como árvores de decisão
 - E semi-supervisionado, como redes bayesianas