

SVM

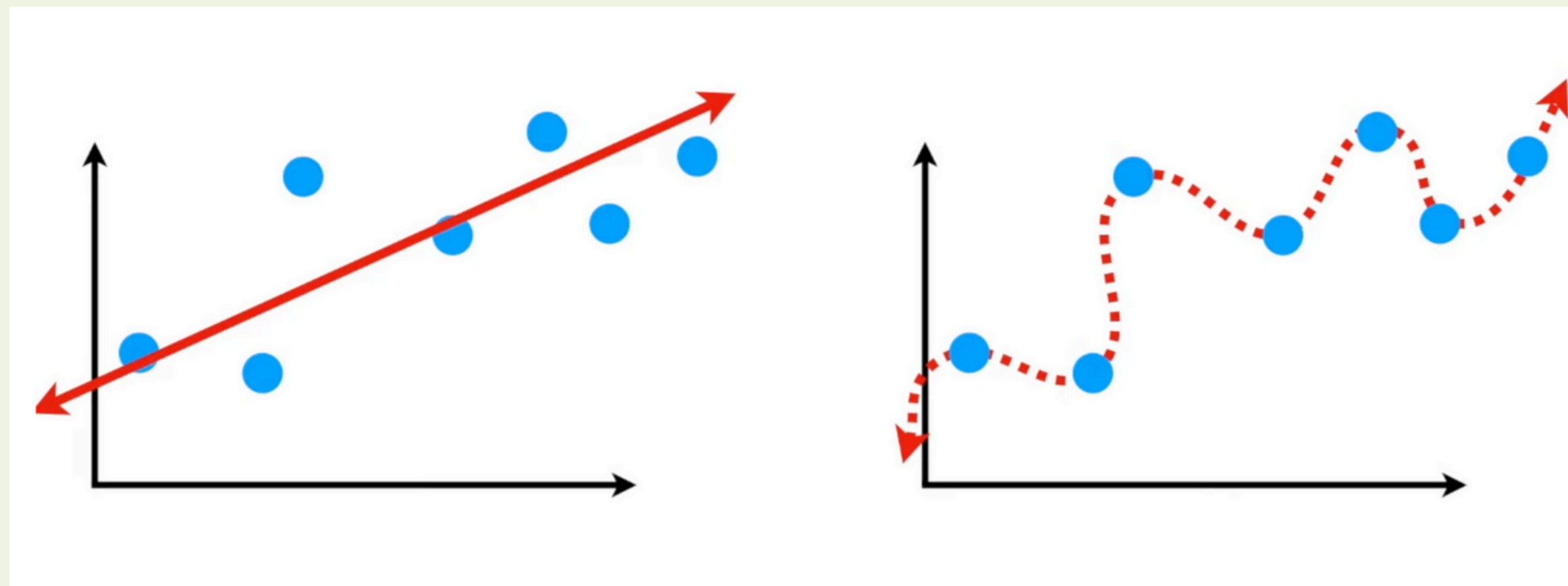
Support Vector Machine
Máquina de Vetores de Suporte

KHÉZIA MOURA
NATÁLIA VIEIRA



UnB
JANEIRO DE 2025

Dilema: Viés/Variância



Se ajusto pouco: risco de underfitting.

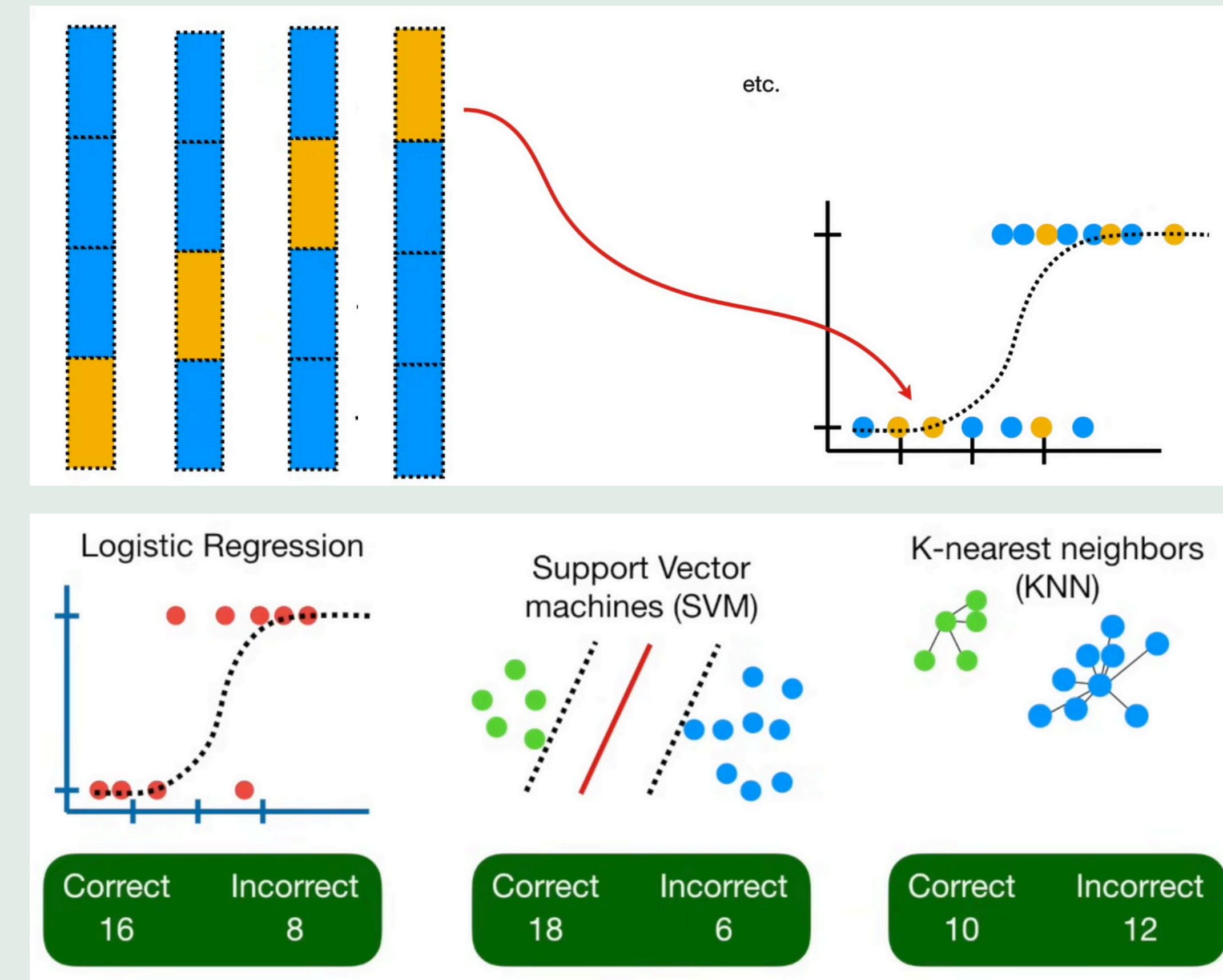
Alto viés, baixa variância.

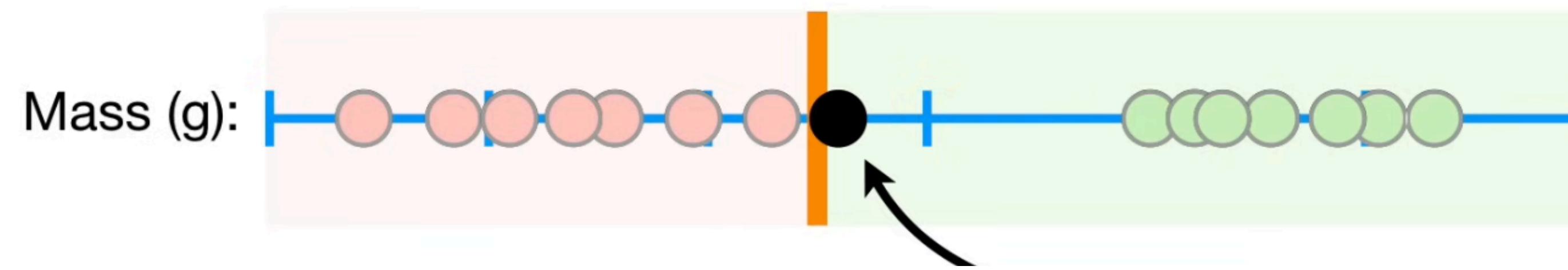
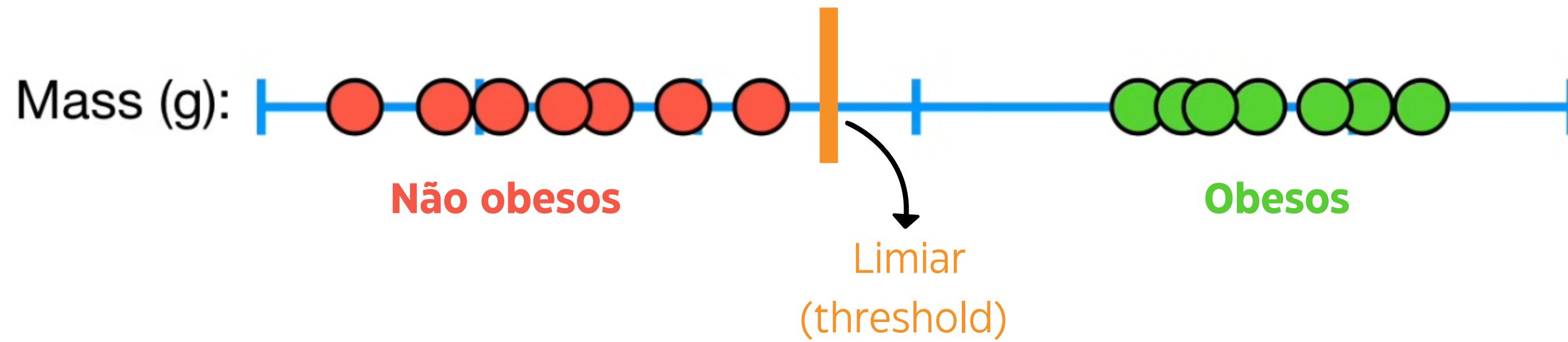
Se ajusto muito: risco de overfitting.

Baixo viés, alta variância no conjunto de teste.

Validação Cruzada

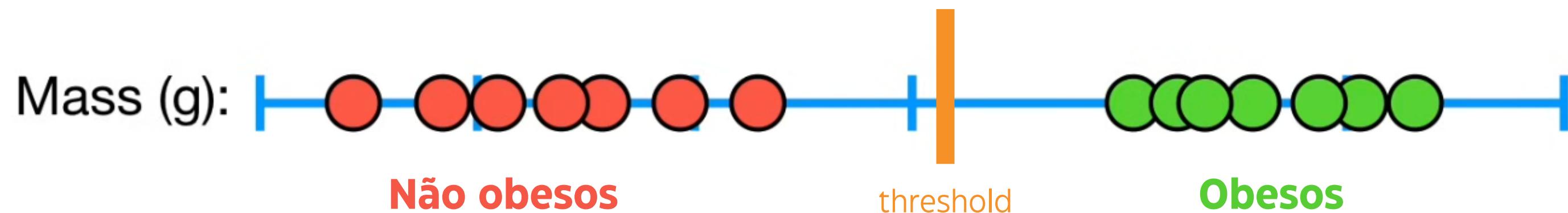
- Treina os modelos em blocos, um de cada vez, combina-os e resume no final, registrando os acertos e erros.
- **Compara os métodos para escolher o que obteve melhor desempenho.**



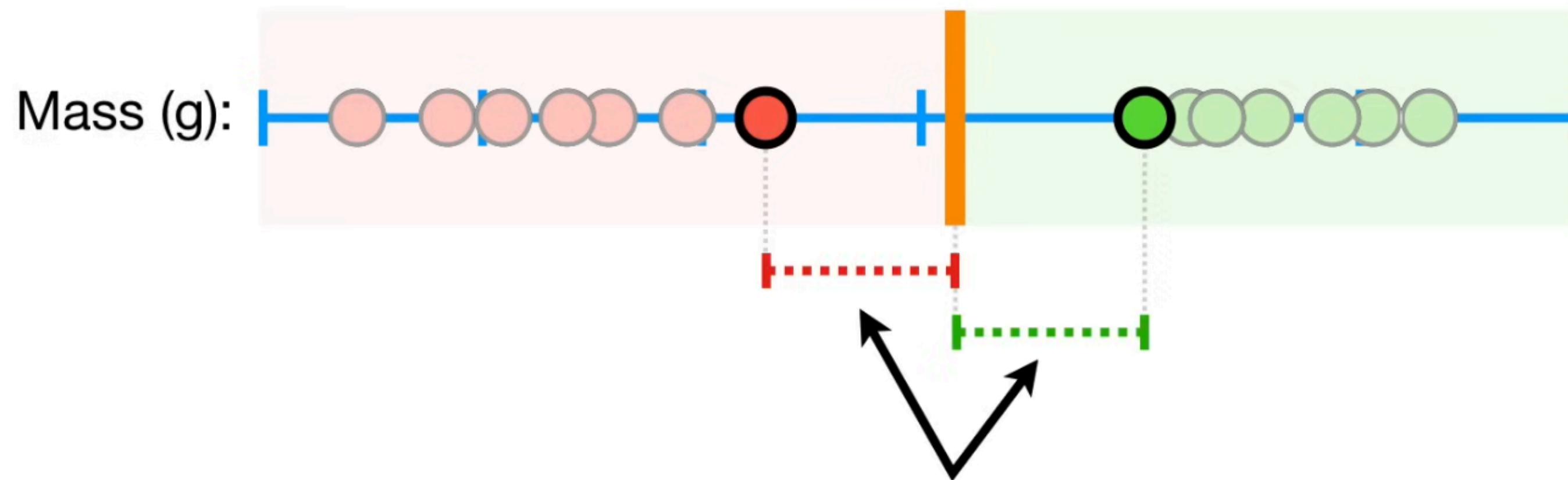


Mais perto dos não obesos, mas classificada
como obesa pela limiar.

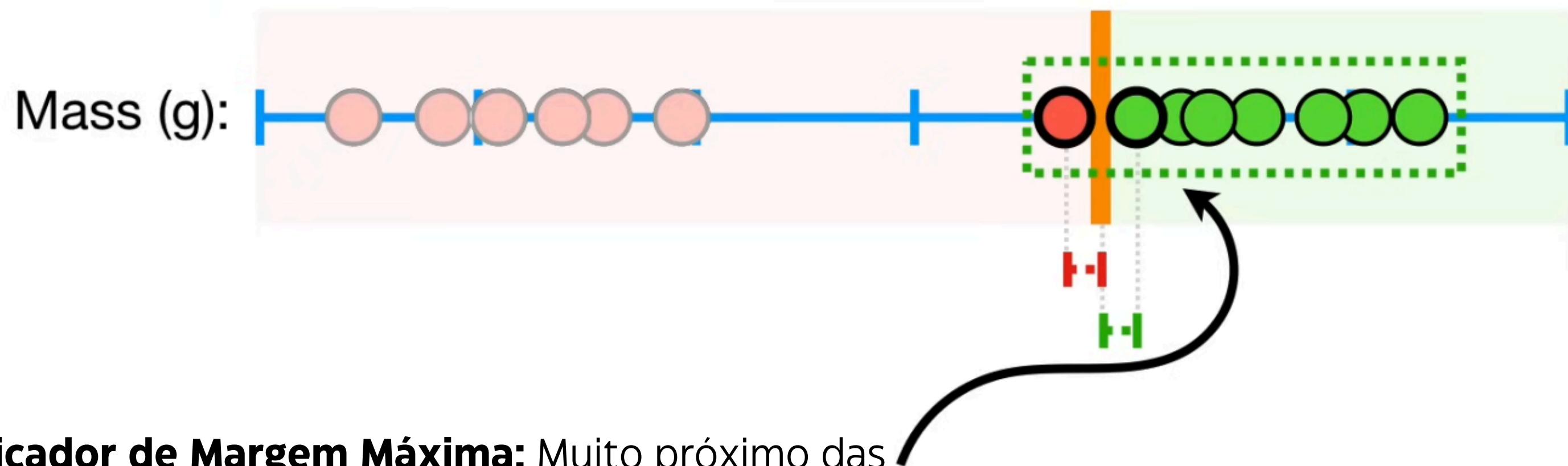
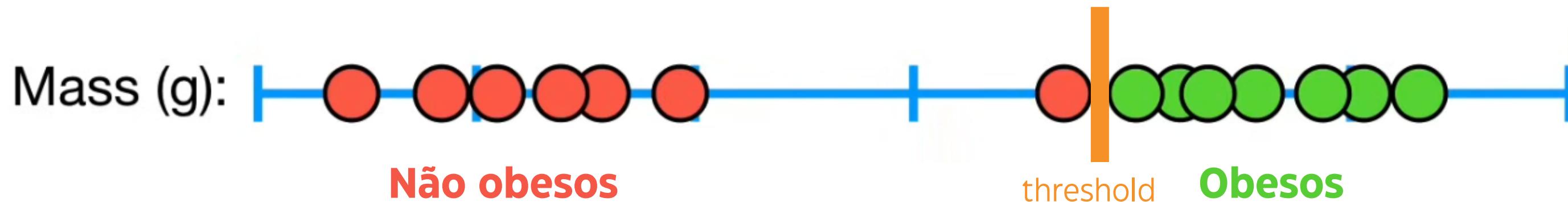
Então o limiar é ruim.



- **Alternativa:** Traçar um limiar que está no meio entre eles.

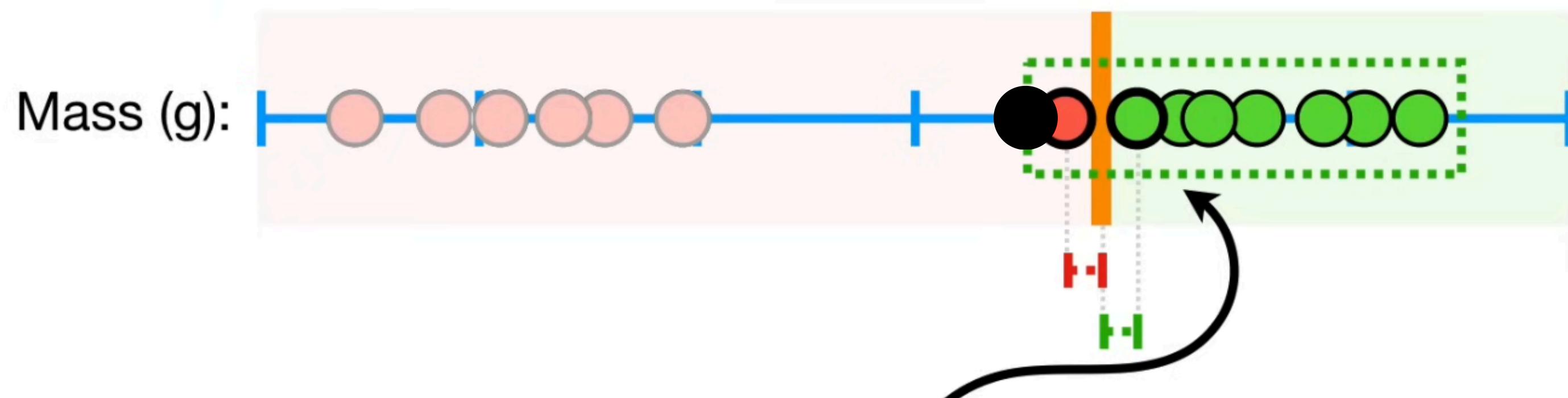
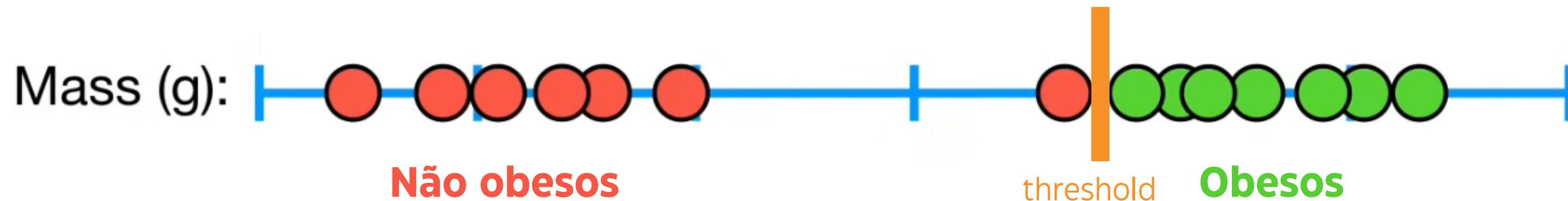


Busca-se maximizar a margem: Classificador de Margem Máxima (Maximal Margin Classifier).



Classificador de Margem Máxima: Muito próximo das observações obesas e longe demais das observações não obesas, pois o máximo da distância entre elas é pequeno.

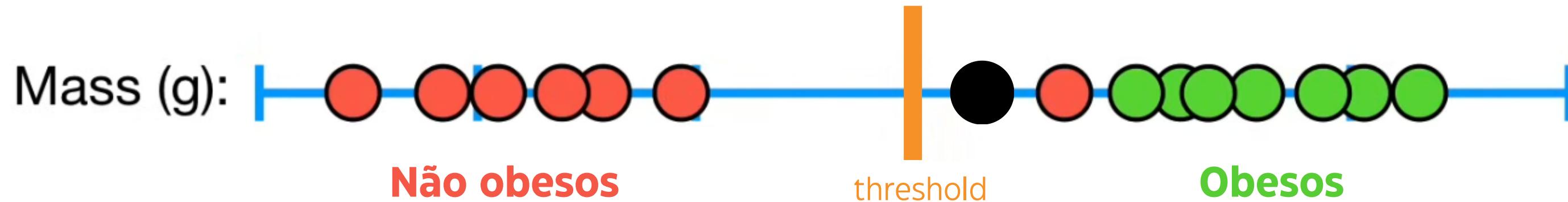
Problema: são sensíveis à valores discrepantes.



Classificador de Margem Máxima: Muito próximo das observações obesas e longe demais das observações não obesas, pois o máximo da distância entre elas é pequeno.

Ao adicionar uma nova observação, ela adota uma classificação obesa embora as observações obesas estejam distantes. Adota uma classificação ruim.

Problema: são sensíveis à valores discrepantes.



Solução: Escolher um limiar que permite erros de classificação (uma margem flexível).

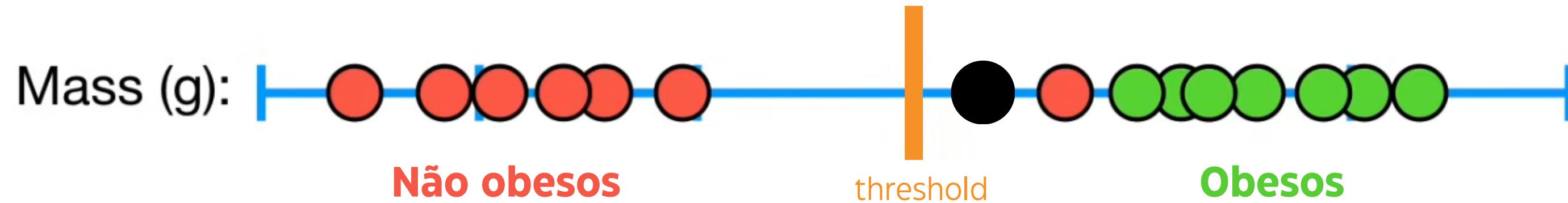
Centralizando a limiar, ainda que 1 obs. seja classificada errada, a próxima será classificada corretamente.

A escolha do limiar entra no dilema viés/variância.

- Limiar sensível aos dados de treinamento: viés baixo, variância alta.
- Limiar menos sensível, que permite o erro de classificação: viés alto, variância pequena.

Como escolher a melhor margem flexível? Validação cruzada.

Para determinar quantos erros de classificação e observação serão permitidas dentro da margem flexível.



Solução: Escolher um limiar que permite erros de classificação (uma margem flexível).

Centralizando a limiar, ainda que 1 obs. seja classificada errada, a próxima será classificada corretamente.

A escolha do limiar entra no dilema viés/variância.

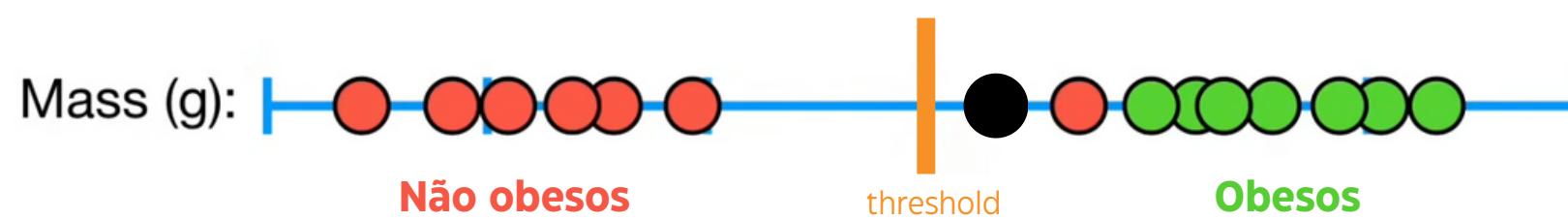
- Limiar sensível aos dados de treinamento: **viés baixo, variância alta**.
- Limiar menos sensível, que permite o erro de classificação: **viés alto, variância pequena**.

Como escolher a melhor margem flexível? Validação cruzada.

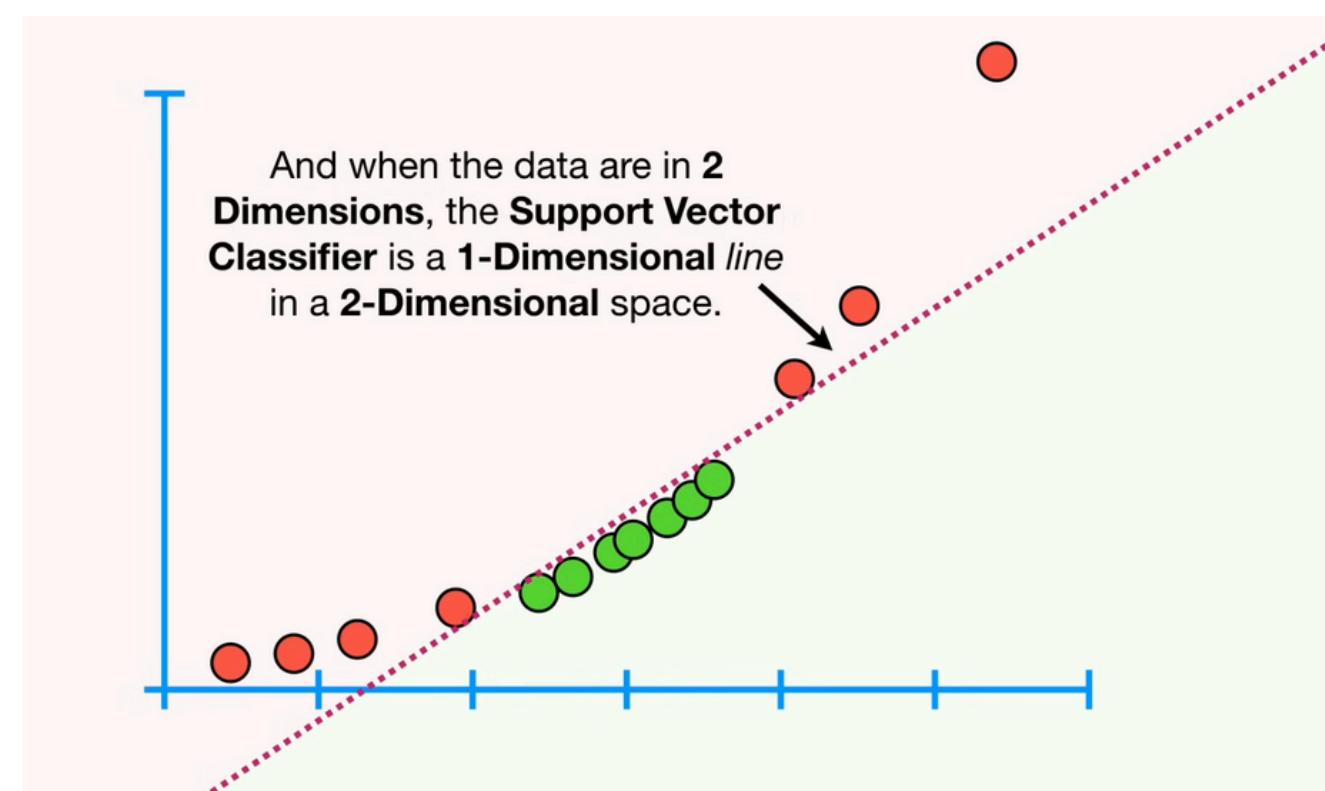
Quando escolhemos uma margem flexível para localizar um limiar, estamos usando um Classificador de Margem Flexível, um **Support Vector Classifier**.

O limiar se comporta de acordo com a dimensão.

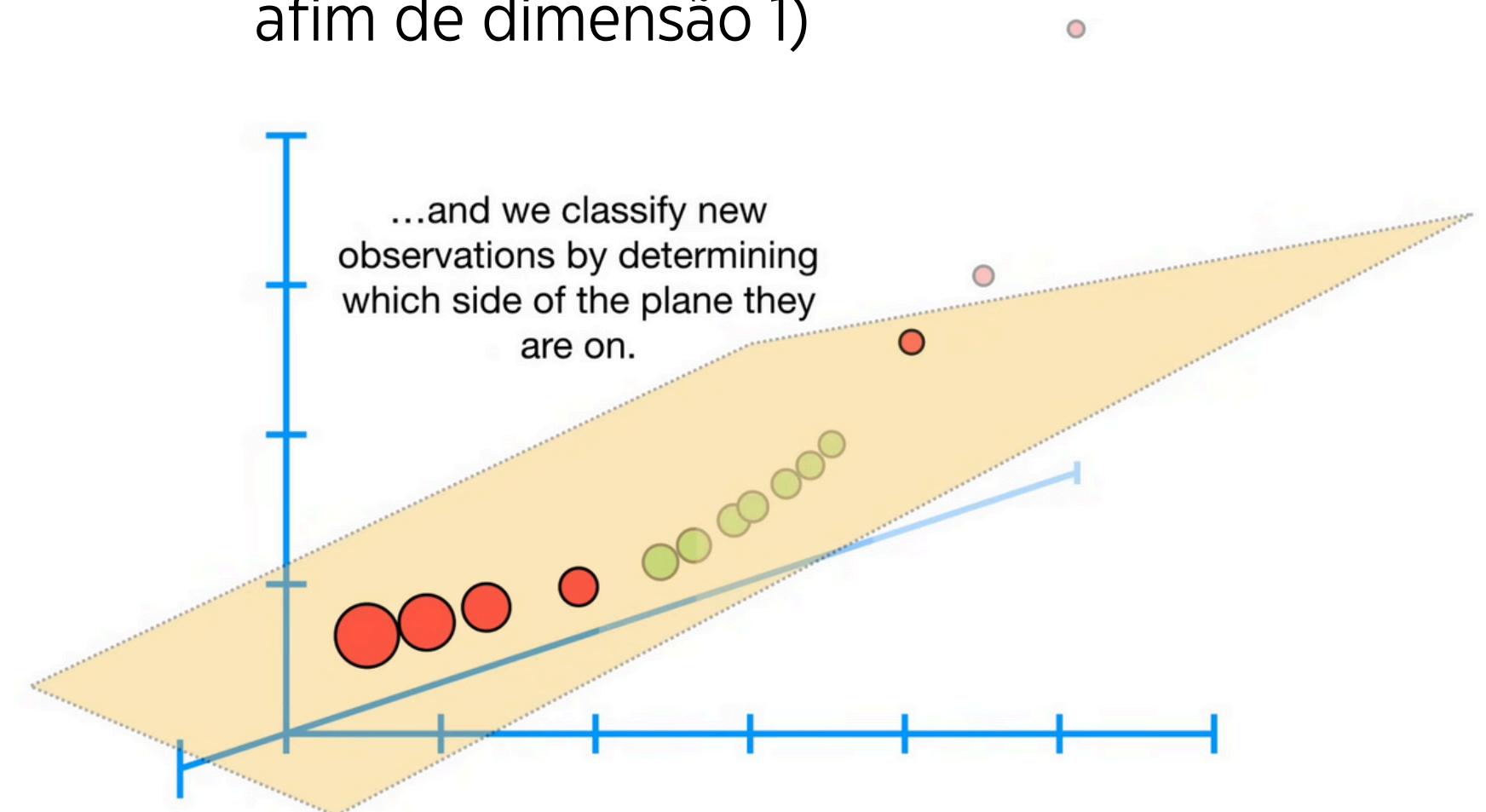
1-D: Ponto (subespaço plano afim de dimensão zero).



2-D: Linha unidimensional (subespaço plano afim de dimensão 1).

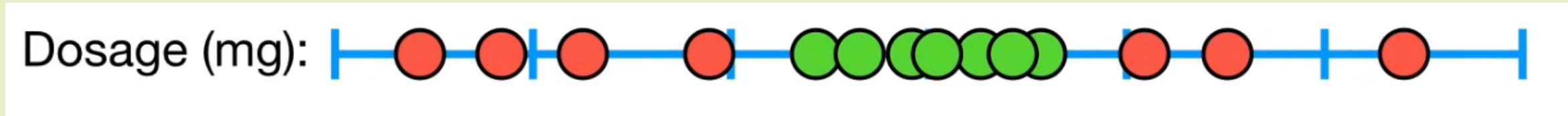


3-D: Plano (subespaço plano afim de dimensão 1)



n-D: Hiperplano (subespaço plano afim).

SVM Não Linear

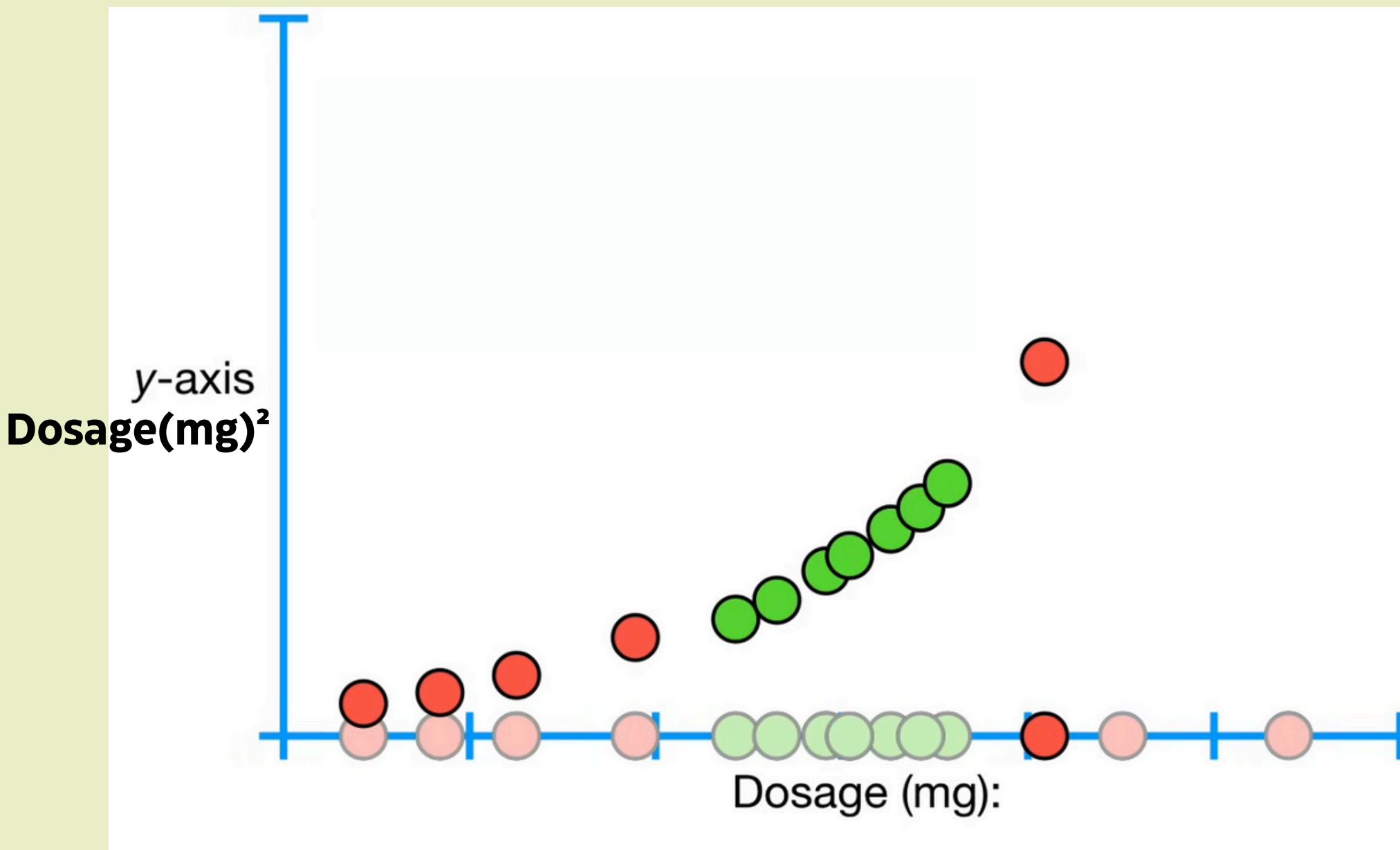
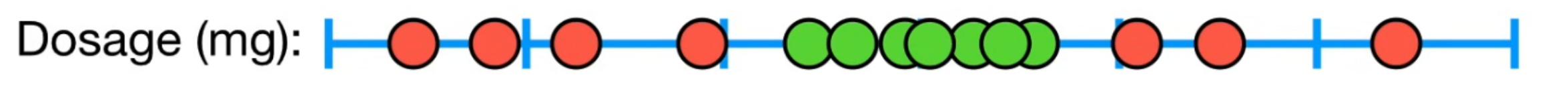


Não importa onde coloquemos o classificador de vetor de suporte, ele vai cometer muitos erros de classificação.

Solução:

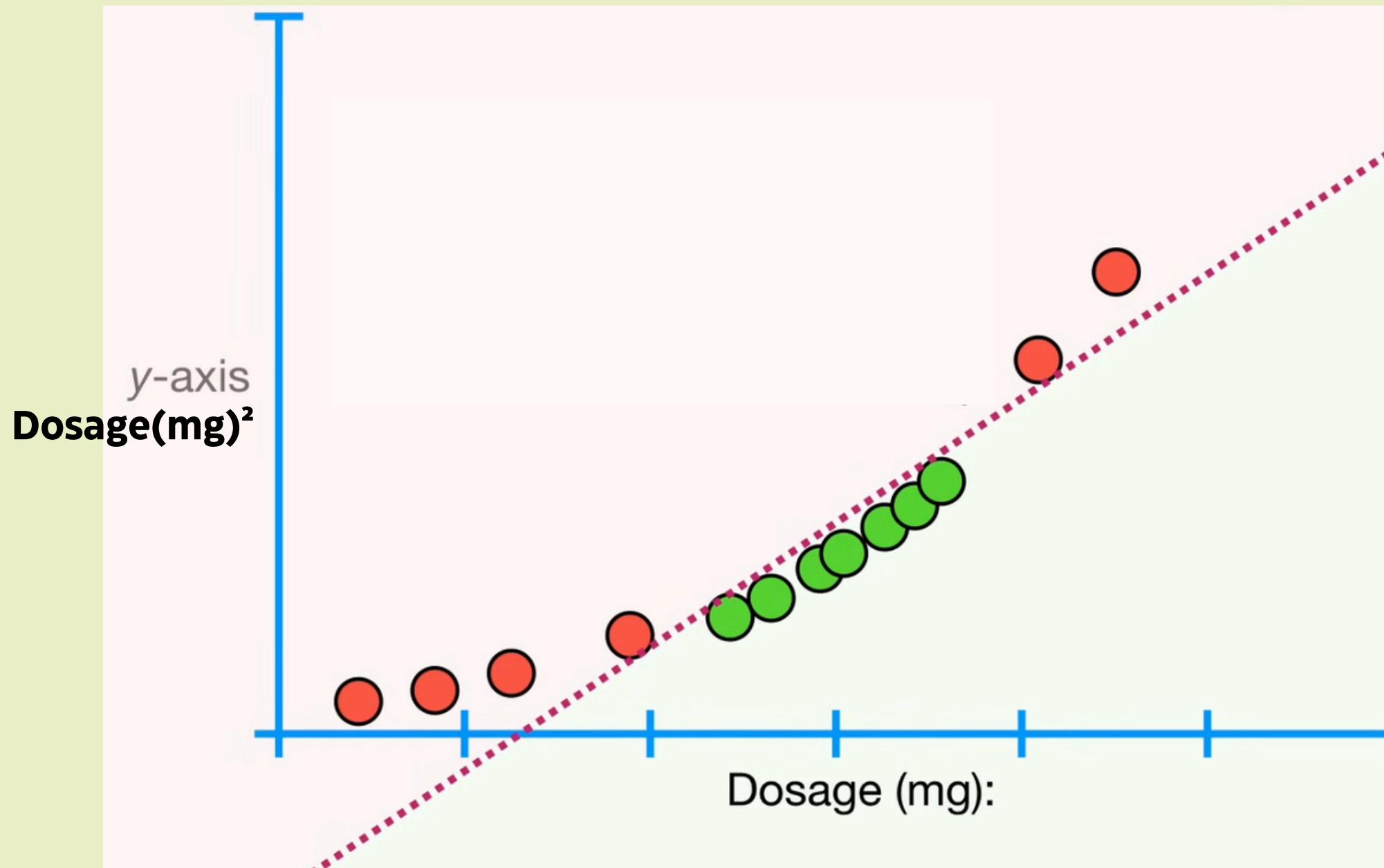
SVM (Support Vector Machine)

SVM Não Linear



SVM Não Linear

Dosage (mg):



Porque não ao cubo?

Porque o SVM utiliza as funções de Kernel que, simplificadamente, encontram as relações entre cada par de observações para definir o limiar.

- Conforme aumento d , melhor capturo essas relações.
- Com a validação cruzada, consigo encontrar o d ideal.

Kernel Polinomial

Polynomial Kernel

$$(a \times b + r)^d$$

Cross Validation

- a e b: observações
- r: coeficiente qualquer
- d: n. da dimensão

Exemplo 1

$$\begin{aligned}(a \times b + \frac{1}{2})^2 &= (a \times b + \frac{1}{2})(a \times b + \frac{1}{2}) \\ &= a^2b^2 + \frac{1}{2}ab + \frac{1}{2}ab + \frac{1}{4}\end{aligned}$$

Expando como
produto e multiplico.

$$= ab + a^2b^2 + \frac{1}{4}$$

Combino e invertendo a
ordem dos termos

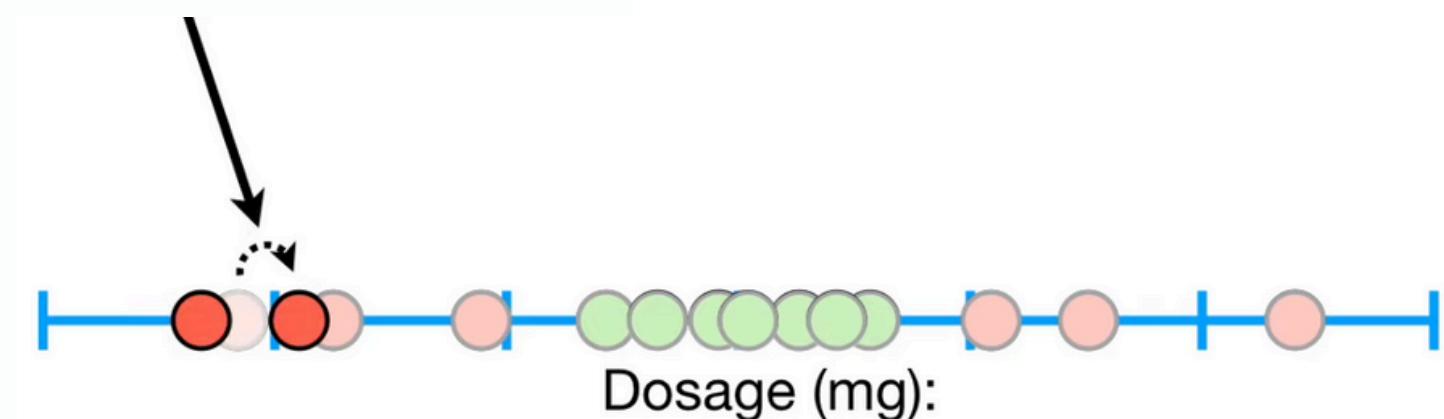
$$= (a, a^2, \frac{1}{2}) \cdot (b, b^2, \frac{1}{2})$$

1º termo 2º termo

Produto escalar

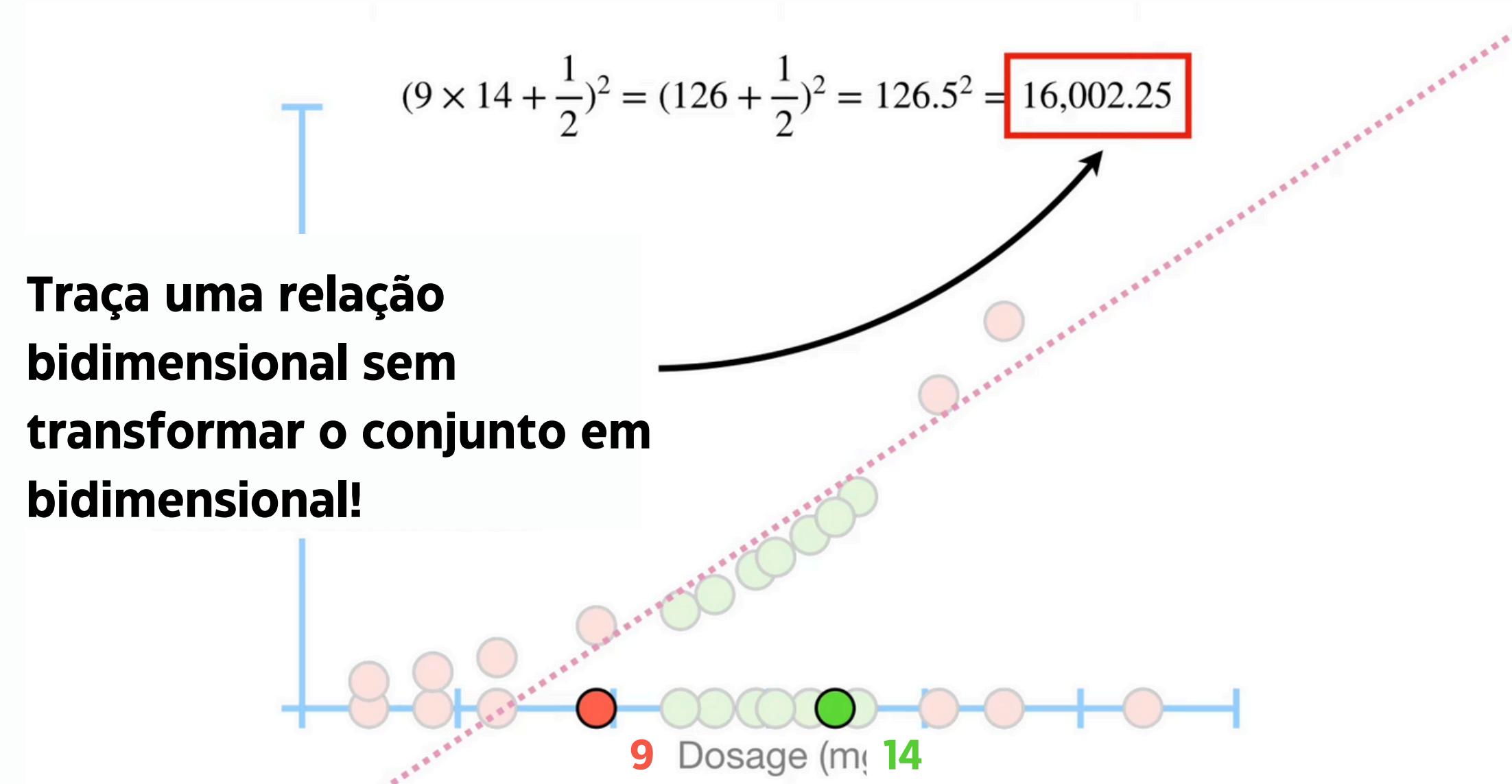
Exemplo 2

$$\begin{aligned}(a \times b + 1)^2 &= (a \times b + 1)(a \times b + 1) \\ &= 2ab + a^2b^2 + 1 \\ &= (\sqrt{2}a, a^2, 1) \cdot (\sqrt{2}b, b^2, 1)\end{aligned}$$



Kernel Polinomial: Porque utilizar?

- **Truque do kernel:** Permite obter o mesmo resultado como se você adicionasse inúmeras características polinomiais, mesmo com polinômios de alto grau, sem realmente precisar adicioná-las. Portanto, não há nenhuma explosão combinatória da quantidade de características, pois não adicionamos nenhuma na verdade.



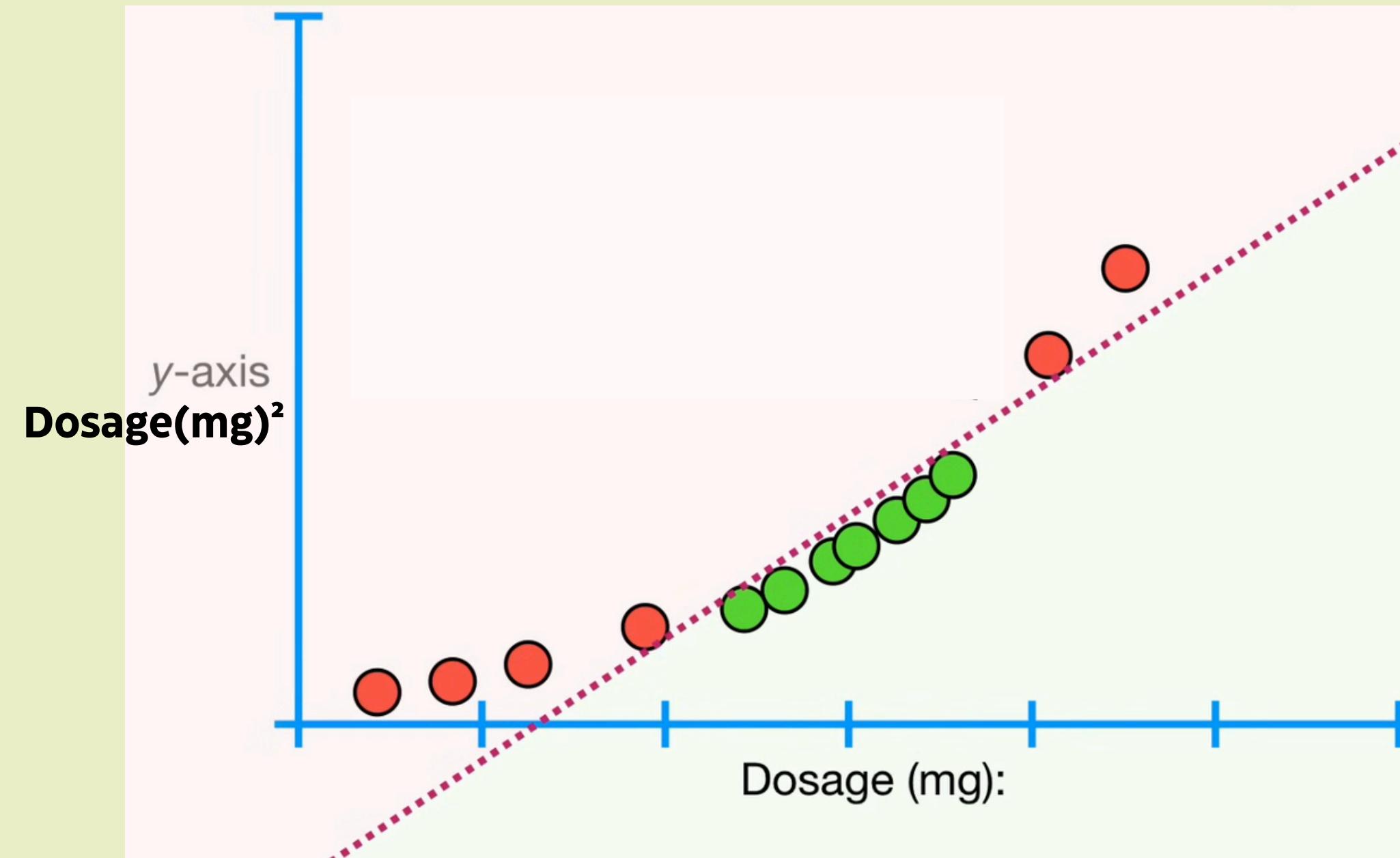
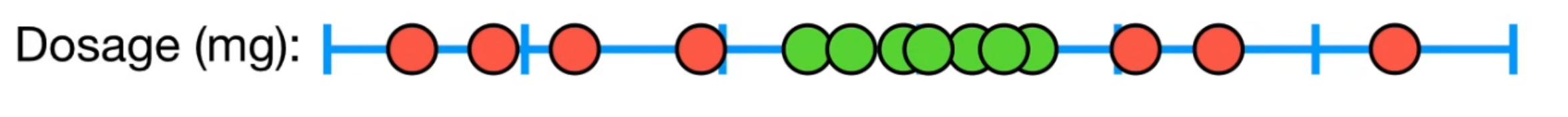
Kernel Polinomial: Porque utilizar?

- **Truque do kernel:** Permite obter o mesmo resultado como se você adicionasse inúmeras características polinomiais, mesmo com polinômios de alto grau, sem realmente precisar adicioná-las. Portanto, não há nenhuma explosão combinatória da quantidade de características, pois não adicionamos nenhuma na verdade. **Traça uma relação bidimensional sem transformar o conjunto em bidimensional!**

Qual Kernel escolher?

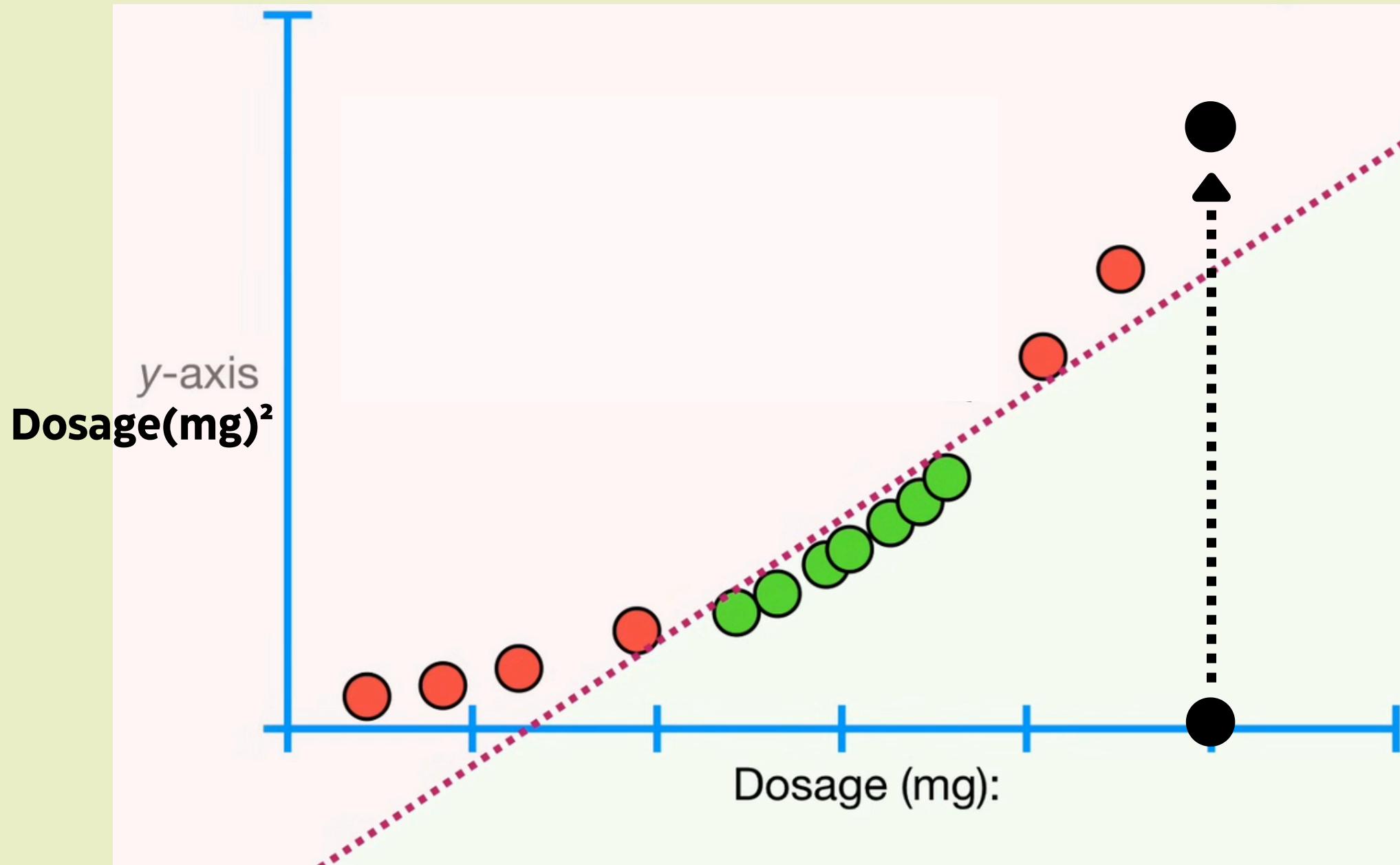
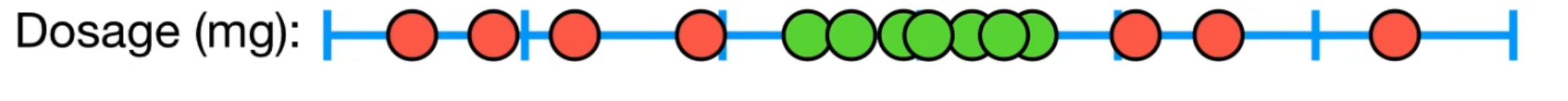
- Primeiro tentar o Kernel Linear, por ser mais rápido.
- Kernel Polinomial;
- Kernel RBF (função de base radial) Gaussiano - que funciona bem na maioria dos casos.
Depende das características dos dados e do caso de uso específico.

SVM



- 1)** Comece com um conj. com dimensão relativamente baixa;
- 2)** Mova-os para uma dimensão maior; **(1D-2D)**
- 3)** Encontre um SVM para classificar.

SVM para previsões

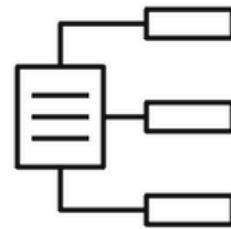


Classificação:

$$\mathbf{w}^T \cdot \mathbf{x} + b = w_1 x_1 + \dots + w_n x_n + b:$$

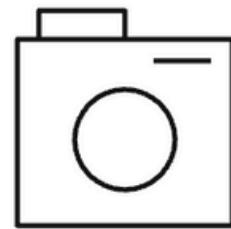
$$\hat{y} = \begin{cases} 0 & \text{se } \mathbf{w}^T \cdot \mathbf{x} + b < 0, \\ 1 & \text{se } \mathbf{w}^T \cdot \mathbf{x} + b \geq 0 \end{cases}$$

Aplicações



Classificação de texto

A SVMs são comumente usadas em processamento de linguagem natural (NLP) para tarefas como análise de sentimento, detecção de spam e modelagem de tópicos. Elas se adequam bem a esses dados porque têm um bom desempenho com dados de alta dimensão.



Classificação de imagens

As SVMs são aplicadas em tarefas de classificação de imagens, como detecção de objetos e recuperação de imagens. Também podem ser úteis em domínios de segurança, classificando uma imagem como adulterada, por exemplo.



Bioinformática

As SVMs também são usadas na classificação de proteínas, análise de expressão gênica e diagnóstico de doenças. As SVMs são frequentemente aplicadas na [pesquisa de câncer](#) (link externo ao site ibm.com) porque podem detectar tendências sutis em conjuntos de dados complexos.



Sistema de informação geográfica (GIS)

As SVMs podem analisar estruturas geofísicas em camadas no subsolo, filtrando o "ruído" dos dados eletromagnéticos. Eles também ajudaram a prever o potencial de liquefação sísmica do solo, o que é relevante para o campo da engenharia civil.

Observações:

- Surge para resolver o problema da Classificação de Margem Suave, que é sensível à outliers,
- O SVM é sensível a escala das variáveis: é preciso escalar previamente.
- Hiperparâmetros: **C** (a flexibilidade da margem), **Kernel** e o **Gama** (o quanto o ponto é influenciável).

Vantagens:

- Útil tanto para classificação e previsão: versabilidade em aplicações.
- Capaz de lidar com dados lineares e não lineares;
- Eficaz em espaços de alta dimensão;

Desvantagens:

- Computacionalmente caro (pode ser lento e demorado);
- Requer escolha de hiperparâmetros.
- Falta de saídas probabilísticas;



Aplicação no R

Fontes

- StatQuest
- Mão à Obra: Aprendizado de Máquina com Scikit-Learn, Keras & TensorFlow (Cap. 5)
- IBM
- <https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107>