CS 145 Discussion 3

Reminders

- HW1 due yesterday 10/19/2017 (Thursday)
- HW2 out today, due Monday Oct 30 ()
- Participation points for Piazza
- Recommendation: Anaconda and PyCharm

HW1 (1)

- $\beta = (X^T X)^{-1} X^T Y$

HW1 (2)

- $L_1 = ?$
- $L_0 = ?$
- Derivatives?

HW1 (2)

- $L_1 = \log(\frac{1}{1+e^x})$
- $L_0 = ?$
- Derivatives?

HW1 (2)

- $L_1 = \log(\frac{1}{1+e^x})$
- $L_0 = \log(\frac{e^x}{1+e^x})$
- Derivatives?

HW1 (3)

- $BR = info([10,10]) - info([9,2],[1,8])$

HW1 (3)

- $BR = info([10,10]) - info([9,2],[1,8]) = 0.397$

Feature Extraction from Real Data

- Types of Features
  - Numerical
  - Categorical
    - Nominal, Binary, Ordinal

- Real data may be messy for extracting features
  - Unorganized structure
  - Hidden and deep information


```json
{
  "created_at": "Tue Nov 24 00:14:03 +0000 2015",
  "id": 668945640508911600,
  "id_str": "668945640508911616",
  "text": "I know that I let you down. Is it too late now to say sorry?",
  "source": "<a href=\"http://twitter.com/download/iphone\" rel=\"nofollow\">Twitter for iPhone</a>",
  "truncated": false,
  "in_reply_to_status_id": null,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null,
  "in_reply_to_user_id_str": null,
  "in_reply_to_screen_name": null,
  "user": {
    "id": 285176507,
    "id_str": "285176507",
    "name": "NINA XELYN",
    "screen_name": "nxelyn",
    "location": "Norway/CA",
    "url": "http://instagram.com/ninaxelyn",
    "description": "sc: nxelyn | gemini | sjsu",
    "protected": false,
    "verified": false,
    "followers_count": 232,
    "friends_count": 64,
    "listed_count": 1,
    "favourites_count": 842,
    "statuses_count": 1332,
    "created_at": "Wed Apr 20 17:28:57 +0000 2011",
    "utc_offset": -28800,
    "time_zone": "Pacific Time (US & Canada)",
    "geo_enabled": true,
    "lang": "en",
    "contributors_enabled": false,
    "is_translator": false,
    "profile_background_color": "C6E2EE",
    "profile_background_image_url": "http://abs.twimg.com/images/themes/theme2/bg.gif",
    "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme2/bg.gif",
    "profile_background_tile": false,
    "profile_link_color": "1F98C7",
    "profile_sidebar_border_color": "C6E2EE",
    "profile_sidebar_fill_color": "DAECF4",
    "profile_text_color": "663B12",
    "profile_use_background_image": true,
    "profile_image_url": "http://pbs.twimg.com/profile_images/668763072245268480/qFxM2HjH_normal.jpg",
    "profile_image_url_https": "https://pbs.twimg.com/profile_images/668763072245268480/qFxM2HjH_normal.jpg",
    "profile_banner_url": "https://pbs.twimg.com/profile_banners/285176507/1444614000",
    "default_profile": false,
    "default_profile_image": false,
    "following": null,
    "follow_request_sent": null,
    "notifications": null
  },
  "geo": null,
  "coordinates": null,
  "place": {
    "id": "7d62cffe6f98f349",
    "url": "https://api.twitter.com/1.1/geo/id/7d62cffe6f98f349.json",
    "place_type": "city",
    "name": "San Jose",
    "full_name": "San Jose, CA",
    "country_code": "US",
    "country": "United States",
    "bounding_box": {
      "type": "Polygon",
      "coordinates": [
        [
          [
            -122.035311,
            37.193164
          ],
          [
            -122.035311,
            37.469154
          ],
          [
            -121.71215,
            37.469154
          ],
          [
            -121.71215,
            37.193164
          ]
        ]
      ]
    },
    "attributes": {}
  },
  "contributors": null,
  "is_quote_status": false,
  "retweet_count": 0,
  "favorite_count": 0,
  "entities": {
    "hashtags": [],
    "urls": [],
    "user_mentions": [],
    "symbols": []
  },
  "favorited": false,
  "retweeted": false,
  "filter_level": "low",
  "lang": "en",
  "timestamp_ms": "1448324043763"
}
```


Example of a tweet data

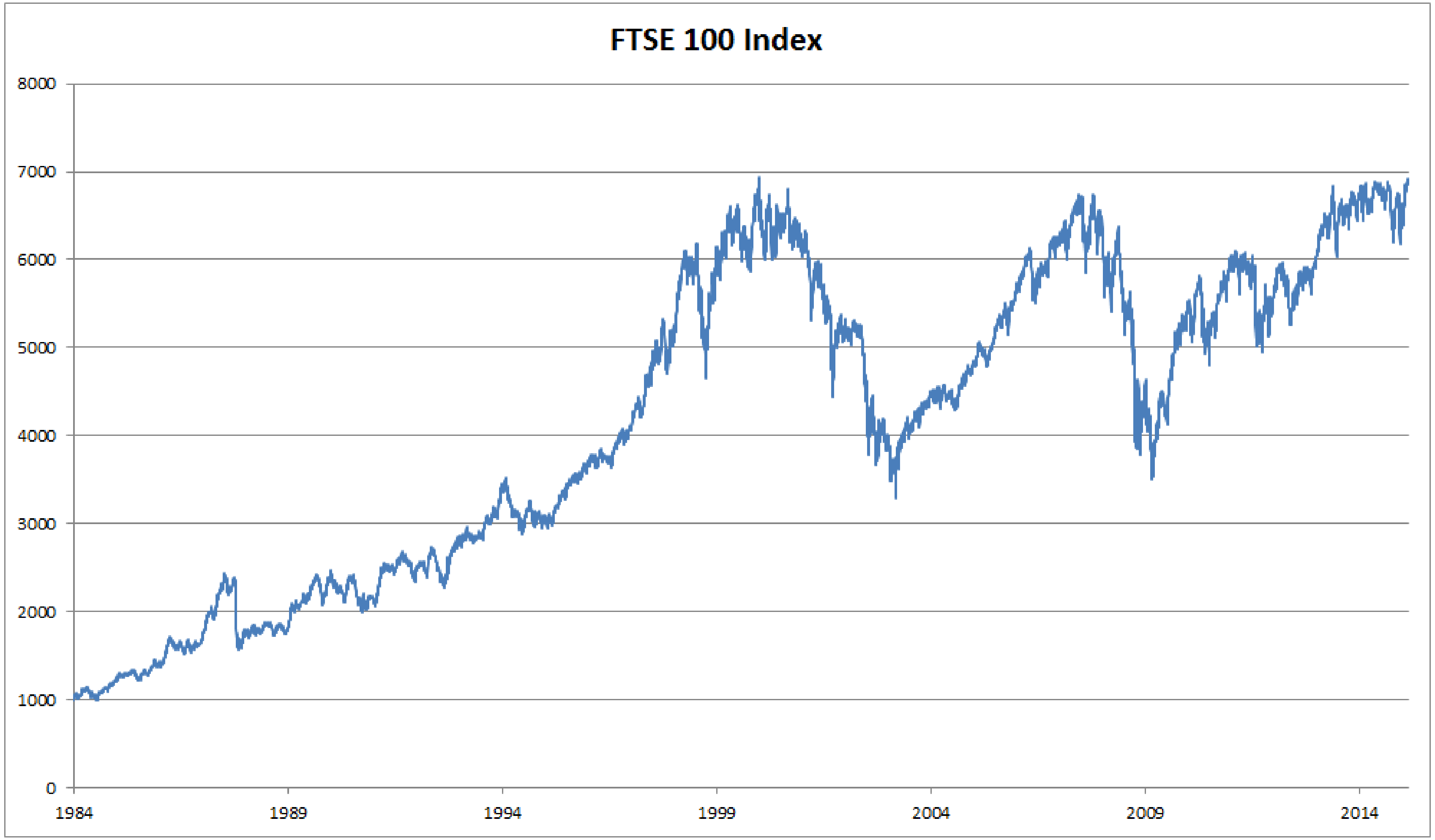
Numerical Features

- Numerical attributes
  - Raw data with numerical formats
  - E.g., numbers of friends and followers, timestamps
- Numerical statistics
  - Numerical statistics towards a characteristic
  - E.g., the length of text, the average daily number of tweets for the user
- Numerical hidden representations
  - Represent data in optimized hidden spaces
  - E.g, pLSA and LDA for text (Week 10)

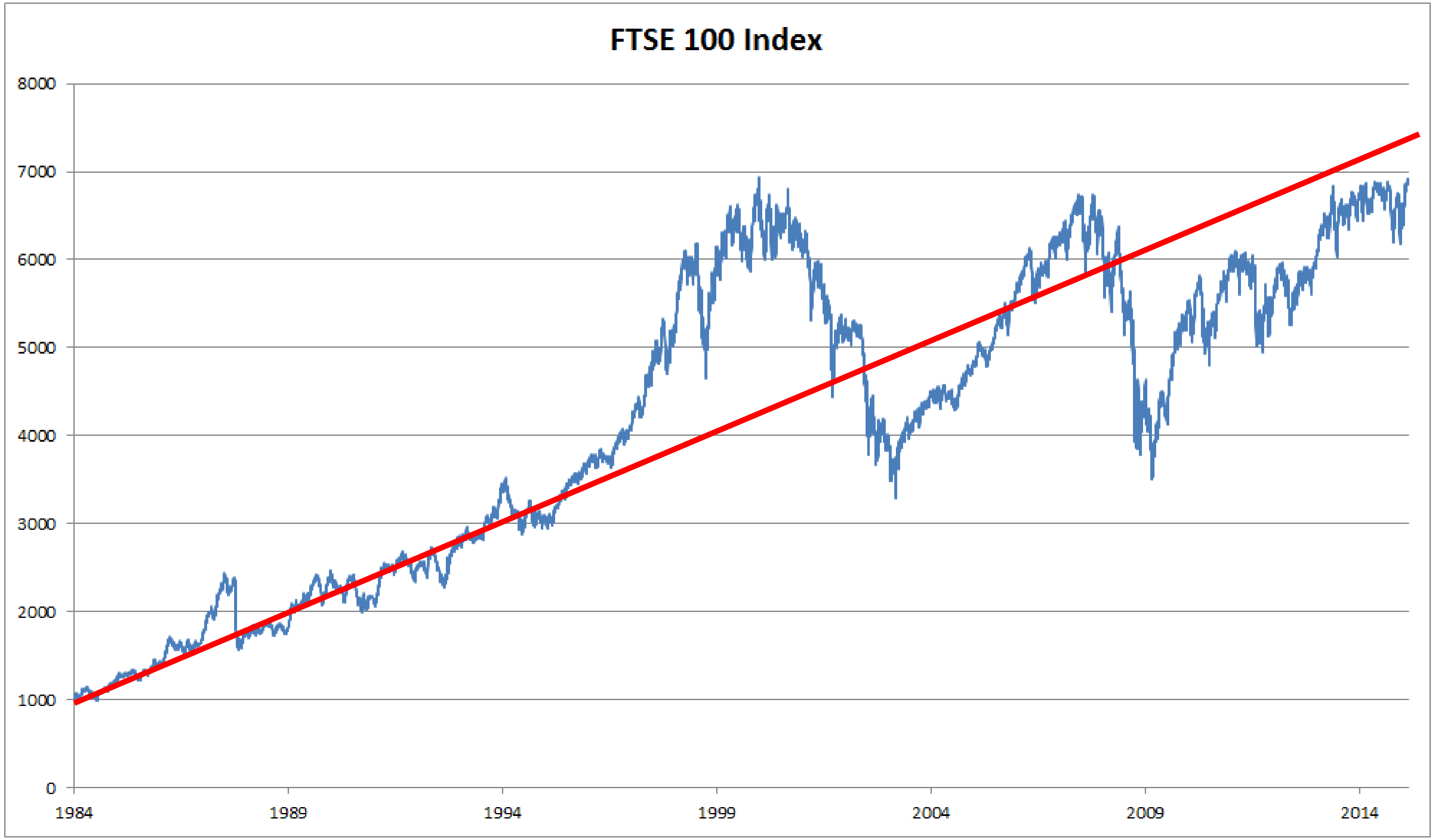
Categorical Features

- Categorical attributes
  - Raw data which originally have a set of discrete categories
  - E.g., cities of users, languages of text,
- Discretization for numerical attributes
  - Transform numerical features into categorical features
  - E.g., Morning/Afternoon/Night, Long/Short Text (more than k words?)
- Categorical statistics
  - Categorical statistics towards a characteristic
  - E.g., If the user posts more than k tweets in a week, Few/Usual/Many tweets posted in near regions

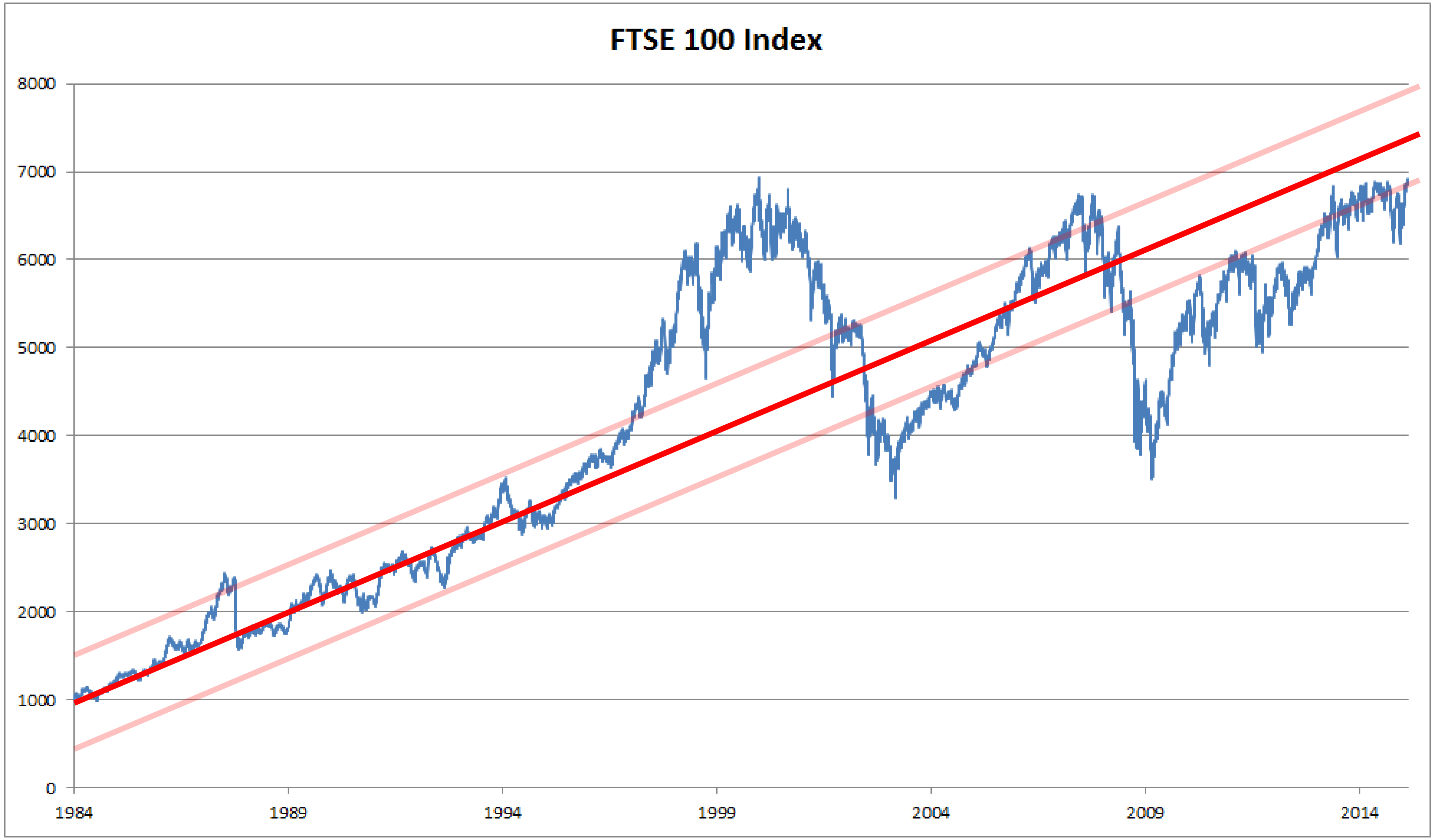
An application of linear regression: Stock Prices



An application of linear regression: Stock Prices



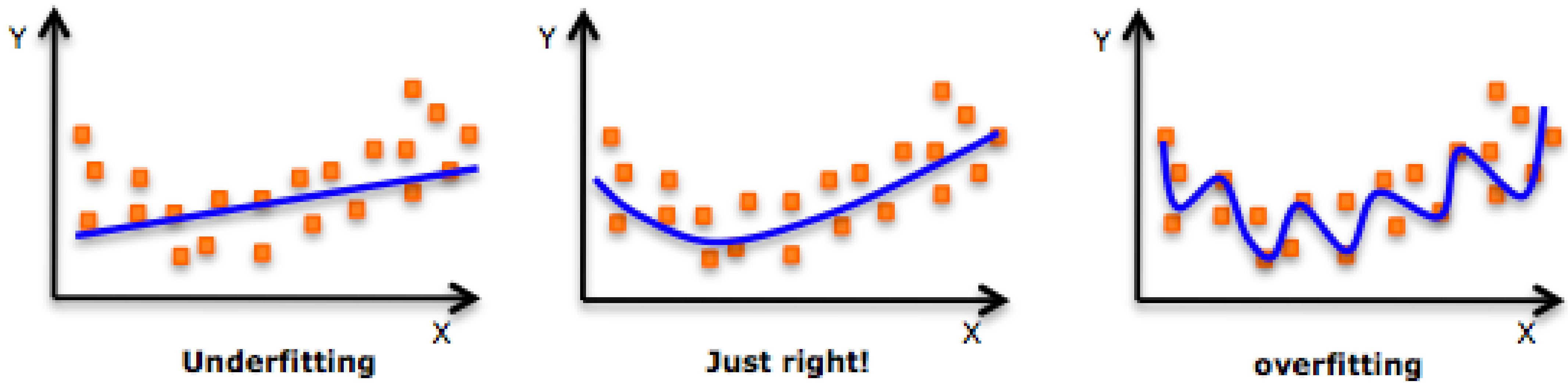
An application of linear regression: Stock Prices



L2 Regularization Derivate

Why Regularization?

- Constraint the parameter values
- Avoid over-fitting phenomenon



L2 Regularization in Linear Regression

- Prediction

$$\hat{y} = \boldsymbol{x}^T\boldsymbol{\beta} = \beta_0 + \sum_i x_i\beta_i$$

- Original Objective

$$\min J(\boldsymbol{\beta}) = \min \sum_{(x,y)} \left(\boldsymbol{x}^T\boldsymbol{\beta} - y\right)^2$$

- L2-Regularized Objective

$$\min \sum_{(x,y)} \left(\boldsymbol{x}^T\boldsymbol{\beta} - y\right)^2 + \lambda||\boldsymbol{\beta}||^2$$

Derivative with L2 Regularization

$$\frac{\partial J(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{(\boldsymbol{x},y)} 2\boldsymbol{x} \cdot \left(\boldsymbol{x}^T \boldsymbol{\beta} - y\right) + \frac{\partial \lambda \|\boldsymbol{\beta}\|^2}{\partial \boldsymbol{\beta}}$$

$$= \sum_{(\boldsymbol{x},y)} 2\boldsymbol{x} \cdot \left(\boldsymbol{x}^T \boldsymbol{\beta} - y\right) + 2\lambda \boldsymbol{\beta}$$

Closed-form with L2-Regularization

$$\nabla_{\boldsymbol{\beta}} J(\boldsymbol{\beta}) = 0$$

$$\nabla_{\boldsymbol{\beta}} \left(\frac{1}{2} ||\boldsymbol{X\beta} - y||^2 + \frac{\lambda}{2} ||\boldsymbol{\beta}||^2 \right) = 0$$

$$\nabla_{\boldsymbol{\beta}} \left(\frac{1}{2} (\boldsymbol{X\beta} - y)^T (\boldsymbol{X\beta} - y) + \frac{\lambda}{2} \boldsymbol{\beta}^T \boldsymbol{\beta} \right) = 0$$

$$\boldsymbol{X}^T \boldsymbol{X\beta} - \boldsymbol{X}^T \boldsymbol{y} + \lambda \boldsymbol{\beta} = 0$$

$$\boldsymbol{X}^T \boldsymbol{X\beta} - \boldsymbol{X}^T \boldsymbol{y} + \lambda \boldsymbol{I\beta} = 0$$

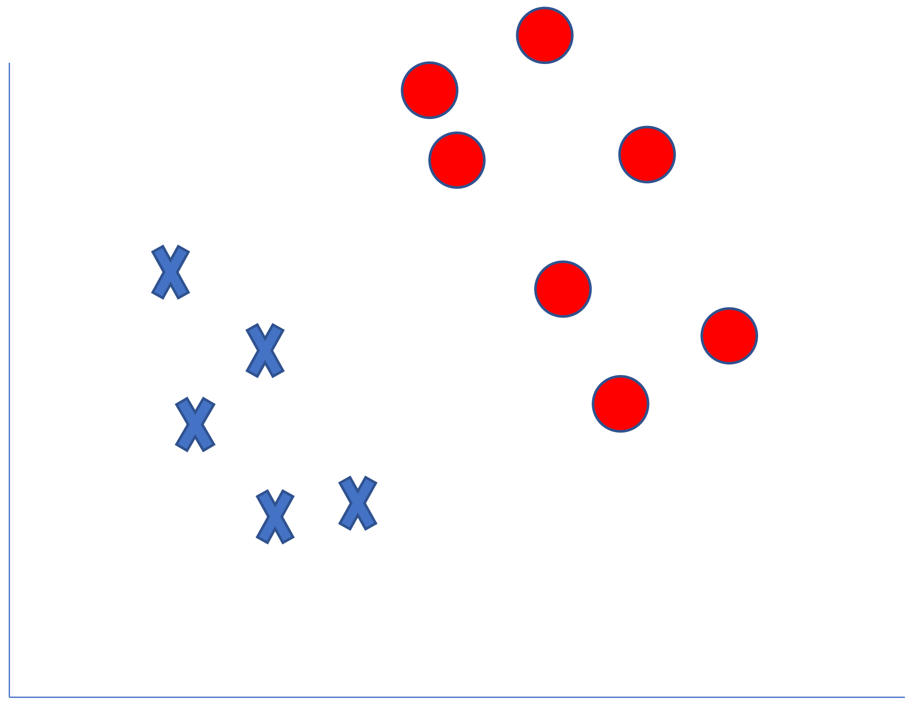
Closed-form with L2-Regularization (Cont'd)

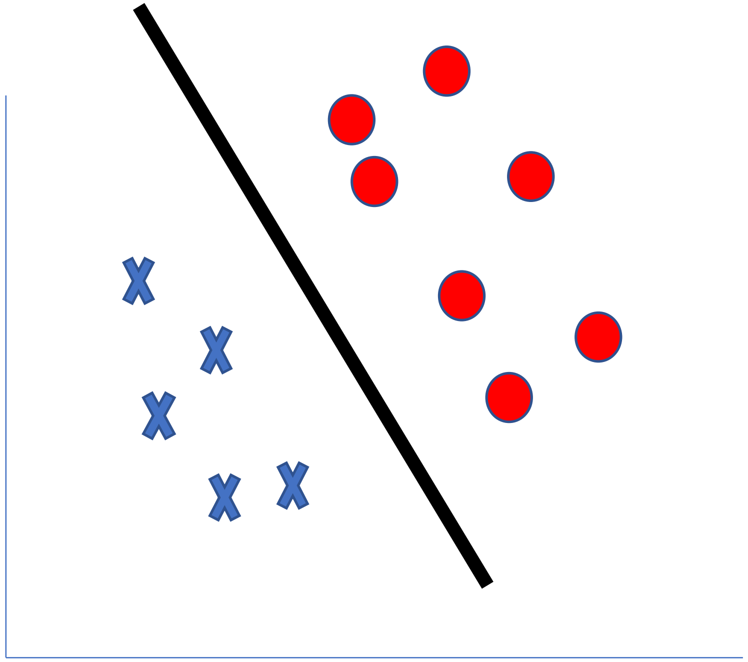
$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{X}^T\boldsymbol{y} + \lambda\boldsymbol{I}\boldsymbol{\beta} = 0$$

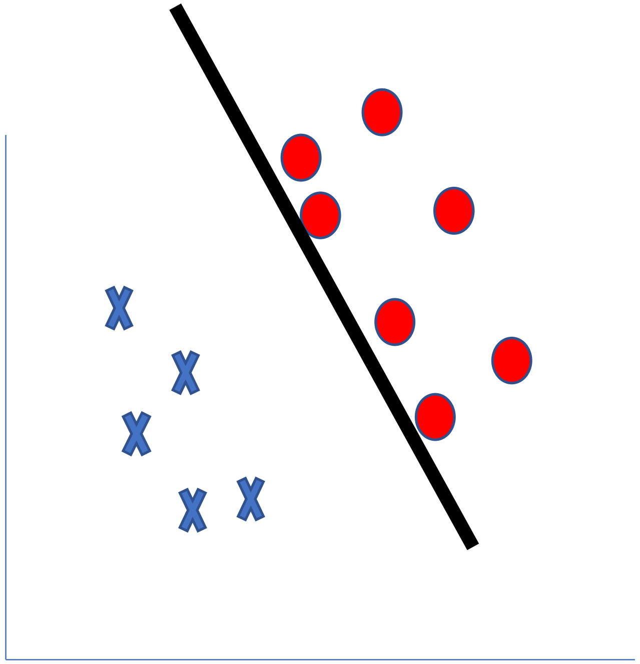
$$\boldsymbol{X}^T\boldsymbol{y} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)\boldsymbol{\beta}$$

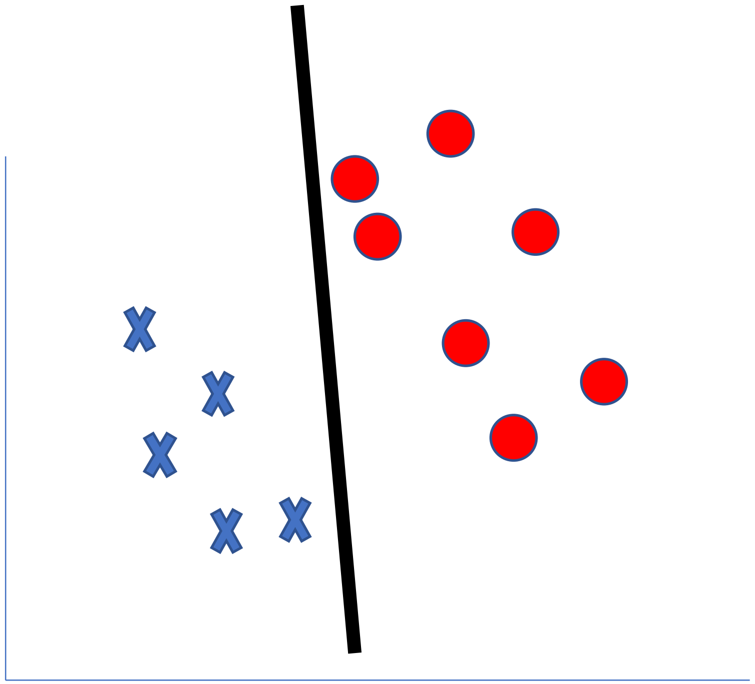
$$\boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Support Vector Machines



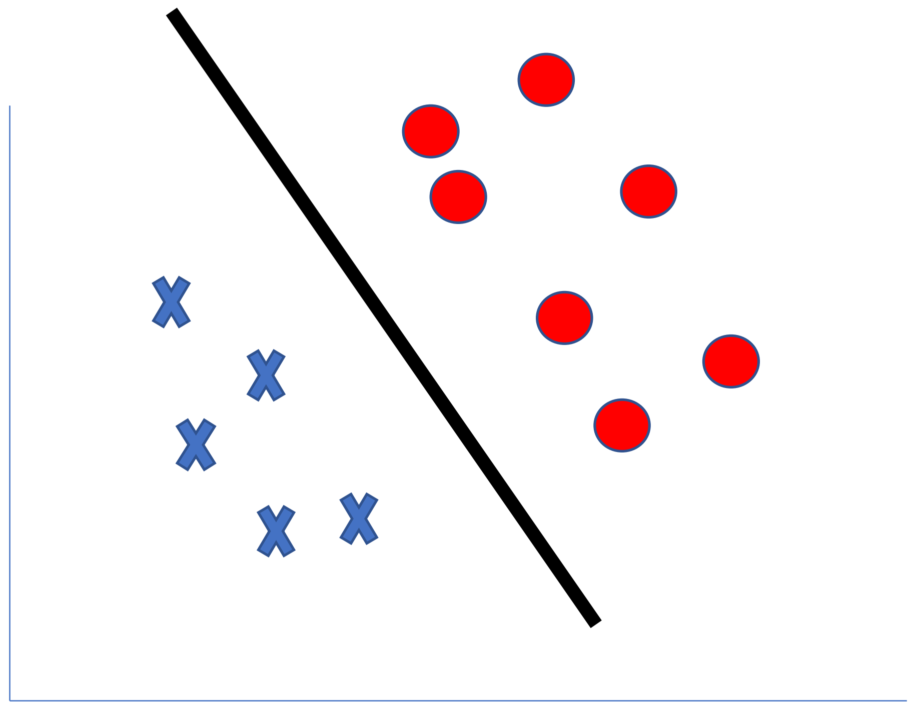






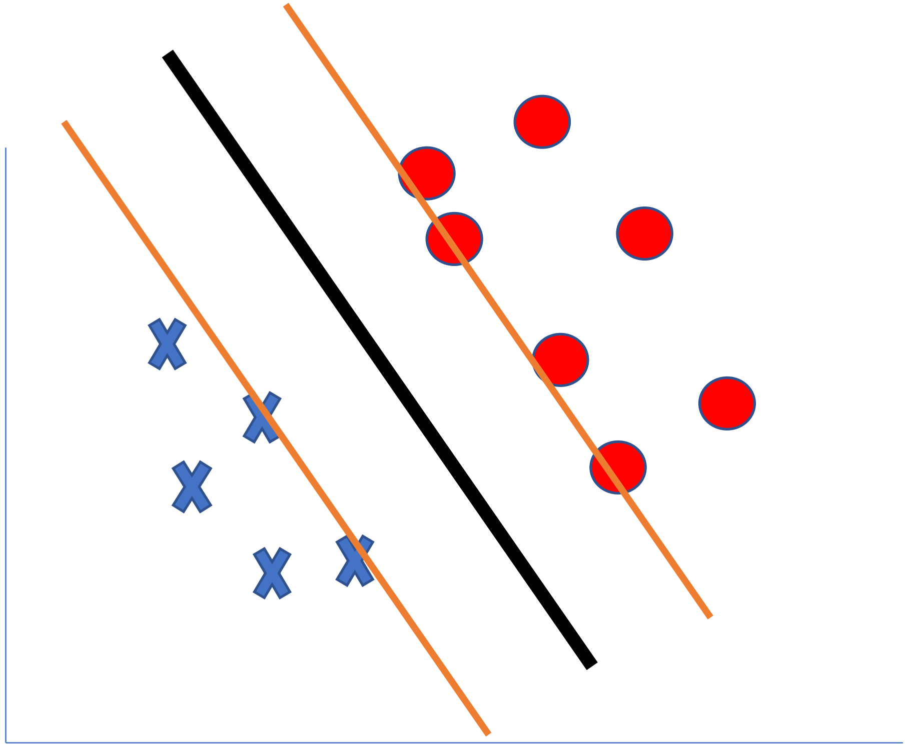
SVM

- What is the margin?



SVM

- What is the margin?



SVM

- How to calculate margin?

SVM

- How to calculate margin?
  - First math!

SVM

- Magnitude of vector?
  - $\vec{v} = < 3,4 >, \left||\vec{v}|\right| = ?$

SVM

- Magnitude of vector?
  - $\vec{v} = < 3,4 >, \left|\left|\vec{v}\right|\right| = 5$

SVM

- Magnitude of vector?
  - $\vec{v} = < 3,4 >, \left|\left|\vec{v}\right|\right| = 5$

SVM

- Direction of vector?
  - $\vec{v} = < 3,4 >?$

SVM

- Direction of vector?
  - $\vec{v} = < 3{,}4 > = direction = < \frac{3}{5}, \frac{4}{5} >$

SVM

- orthogonal projection of $\vec{x}$ onto $\vec{y}$?
  - $\vec{x} = < 3,4 >, \vec{y} = < 2, 1 >$

SVM

- orthogonal projection of $\vec{x}$ onto $\vec{y}$?
    - $\vec{x} = < 3,4 >, \vec{y} = < 2, 1 >$
    - $\left(\frac{\vec{y}}{||\vec{y}||} \cdot \vec{x}\right) \times \frac{\vec{y}}{||\vec{y}||}$

SVM

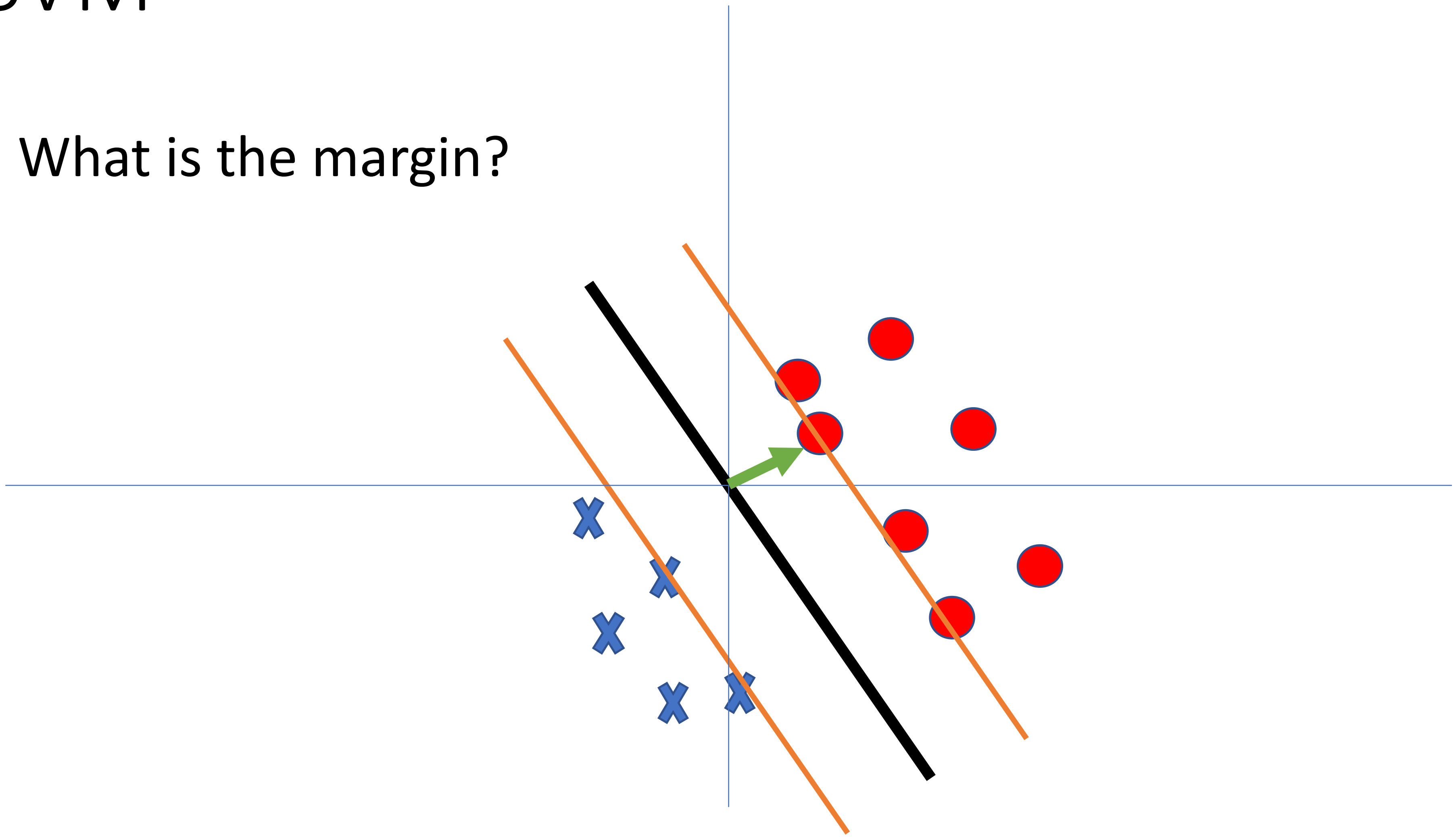
- orthogonal projection of $\vec{x}$ onto $\vec{y}$?
  - $\vec{x} = < 3{,}4 >, \vec{y} = < 2, 1 > = \left(< \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > \cdot < 3{,}4 >\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} >$
  - $\left(\frac{\vec{y}}{||\vec{y}||} \cdot \vec{x}\right) \times \frac{\vec{y}}{||\vec{y}||}$

SVM

- orthogonal projection of $\vec{x}$ onto $\vec{y}$?
  - $\vec{x} = < 3{,}4 >, \vec{y} = < 2, 1 > = \left(< \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > \cdot < 3{,}4 >\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} >$
  - $= \left(\frac{6}{\sqrt{5}} + \frac{4}{\sqrt{5}}\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > = < 4{,}2 >$

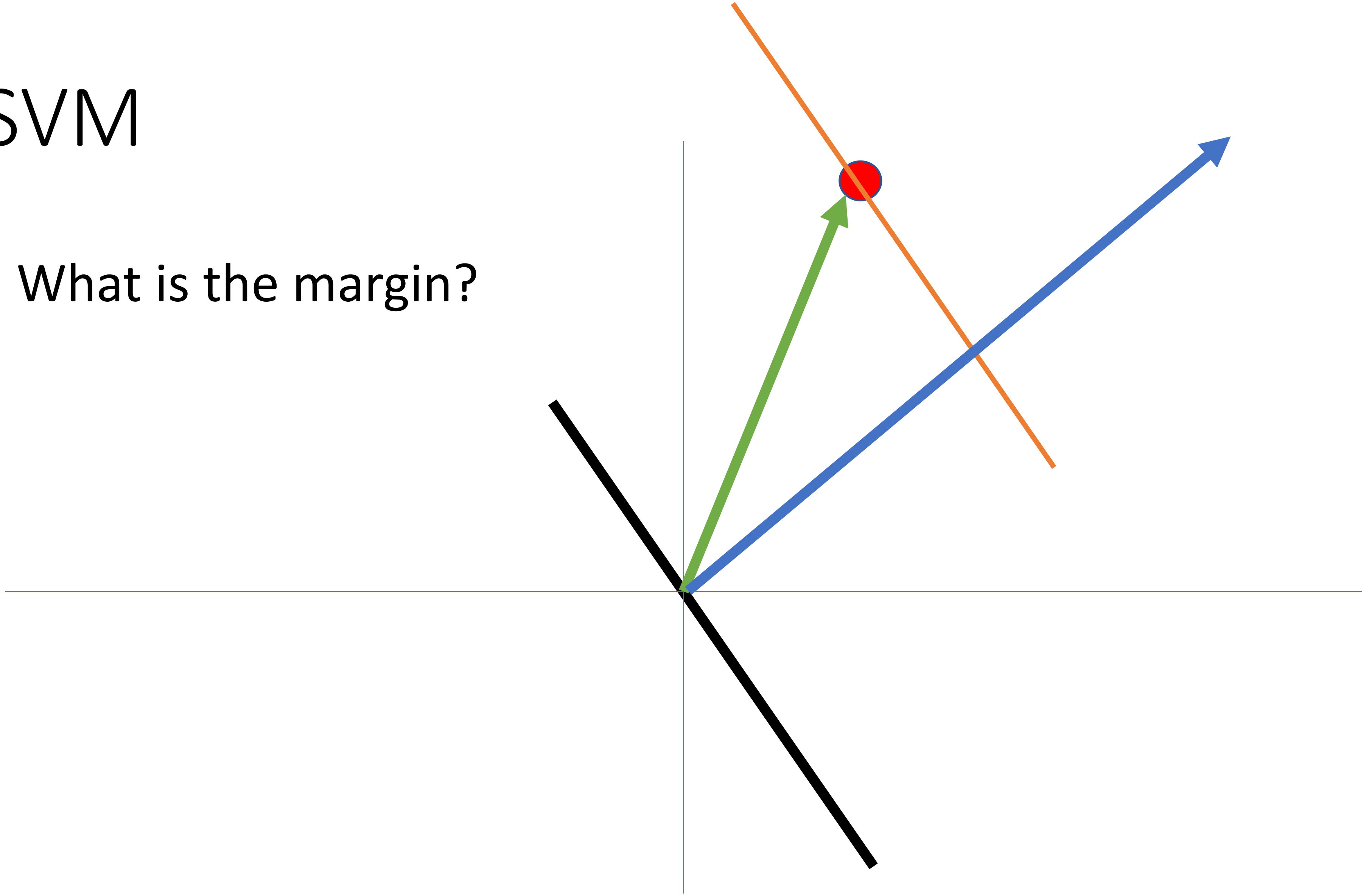
SVM

- What is the margin?



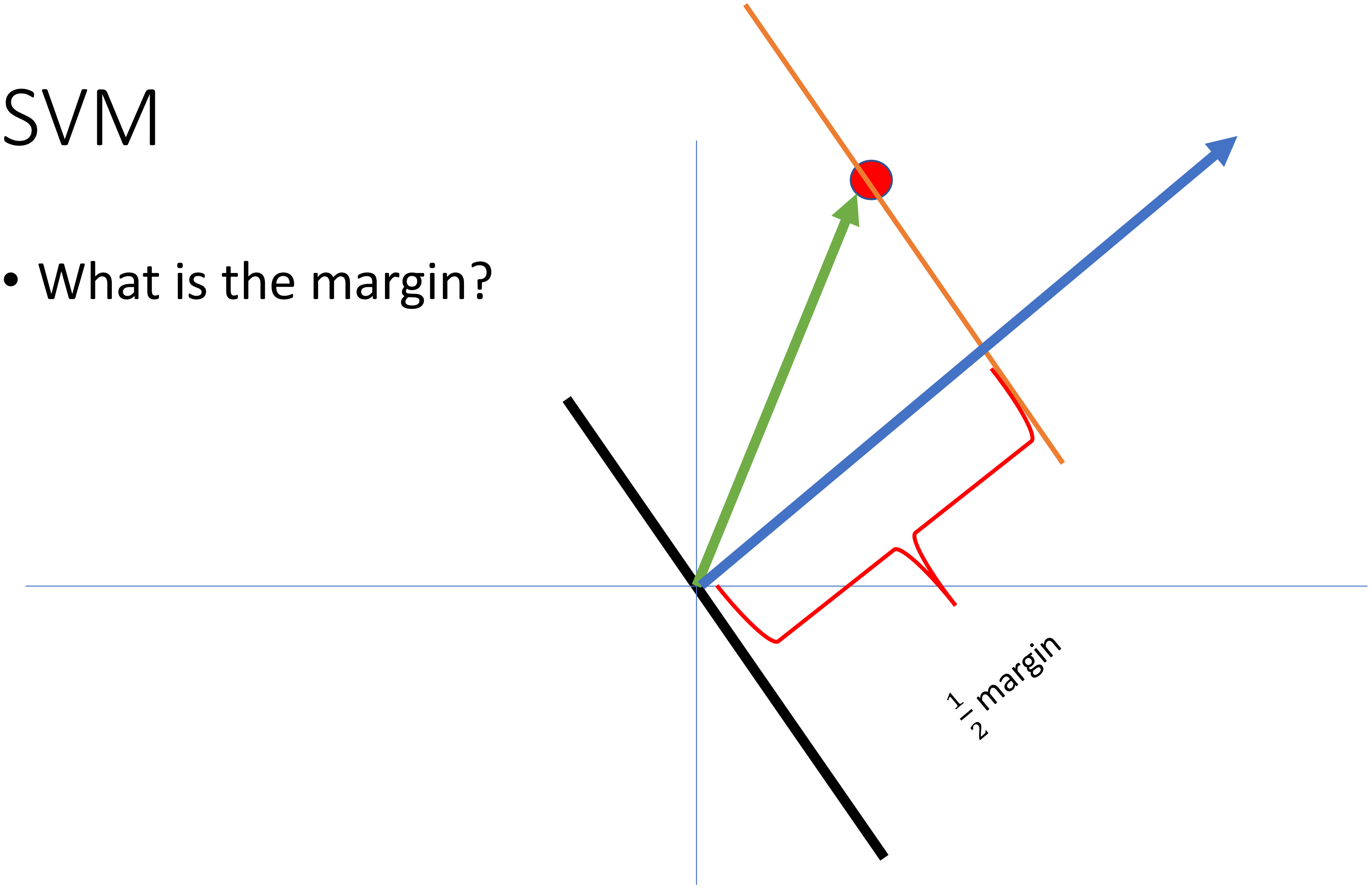
SVM

- What is the margin?



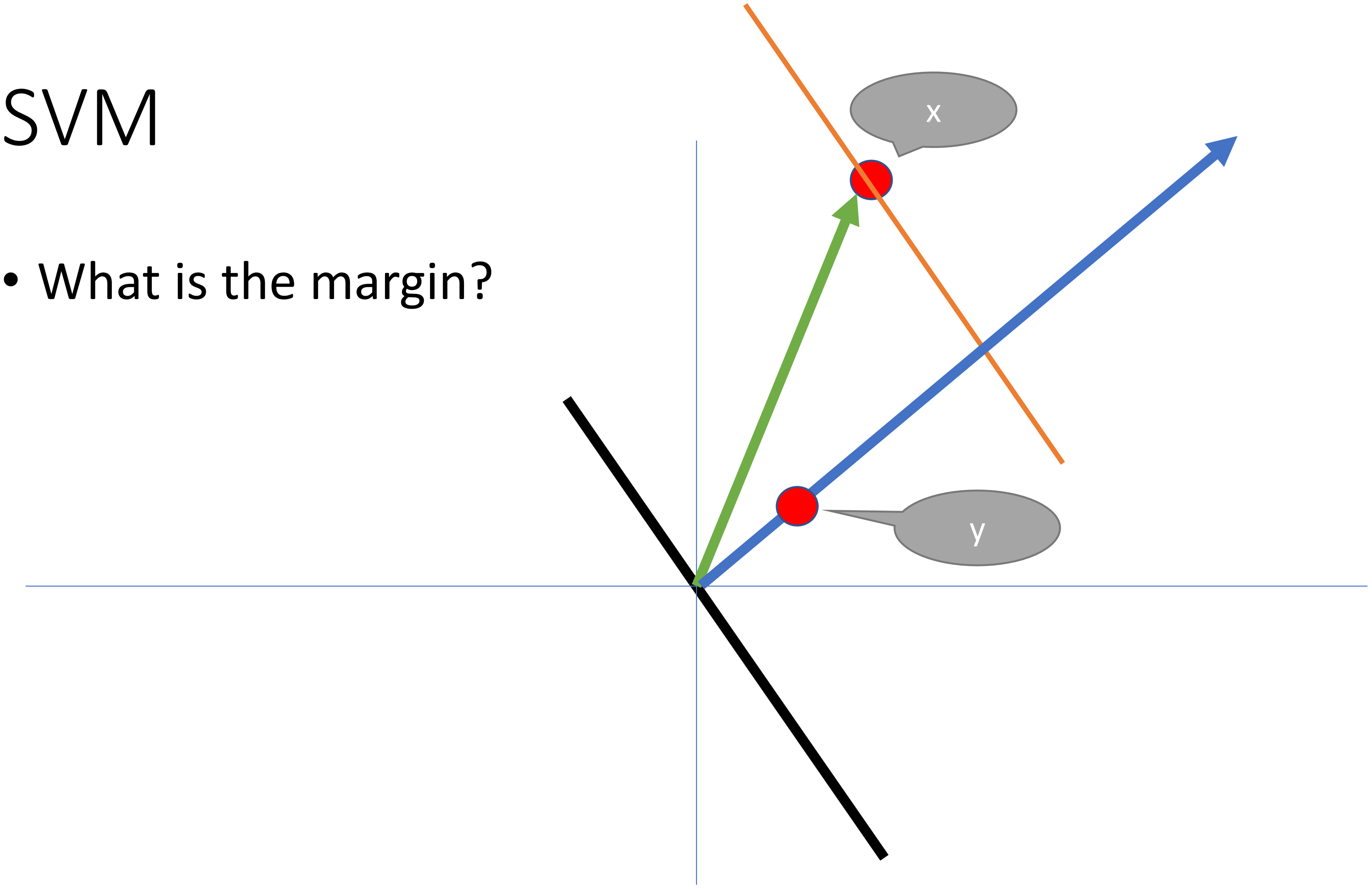
SVM

- What is the margin?



SVM

- What is the margin?



SVM

- What is the margin?
  - $\vec{x} = < 3,4 >, \vec{y} = < 2, 1 >$
- orthogonal projection of $\vec{x}$ onto $\vec{y}$?
  - $\vec{x} = < 3,4 >, \vec{y} = < 2, 1 > = \left(< \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > \cdot < 3,4 >\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} >$
  - $= \left(\frac{6}{\sqrt{5}} + \frac{4}{\sqrt{5}}\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > = < 4,2 > = \vec{p}$

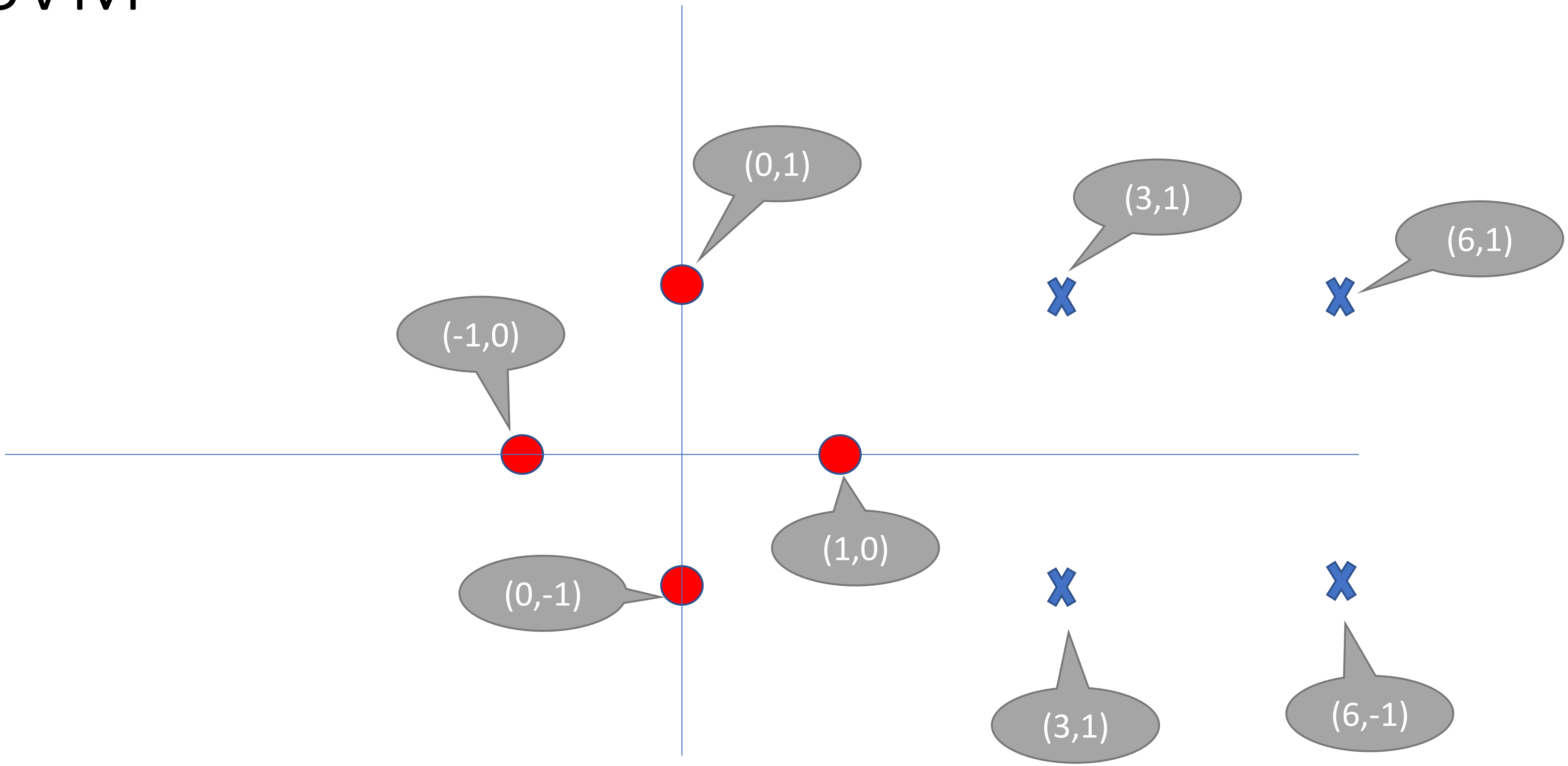
SVM

- What is the margin?
  - $\vec{x} = < 3,4 >, \vec{y} = < 2, 1 > = 2 \cdot ||\vec{p}||$
- orthogonal projection of $\vec{x}$ onto $\vec{y}$?
  - $\vec{x} = < 3,4 >, \vec{y} = < 2, 1 > = \left(< \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > \cdot < 3,4 >\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} >$
  - $= \left(\frac{6}{\sqrt{5}} + \frac{4}{\sqrt{5}}\right) \times < \frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}} > = < 4,2 > = \vec{p}$

SVM

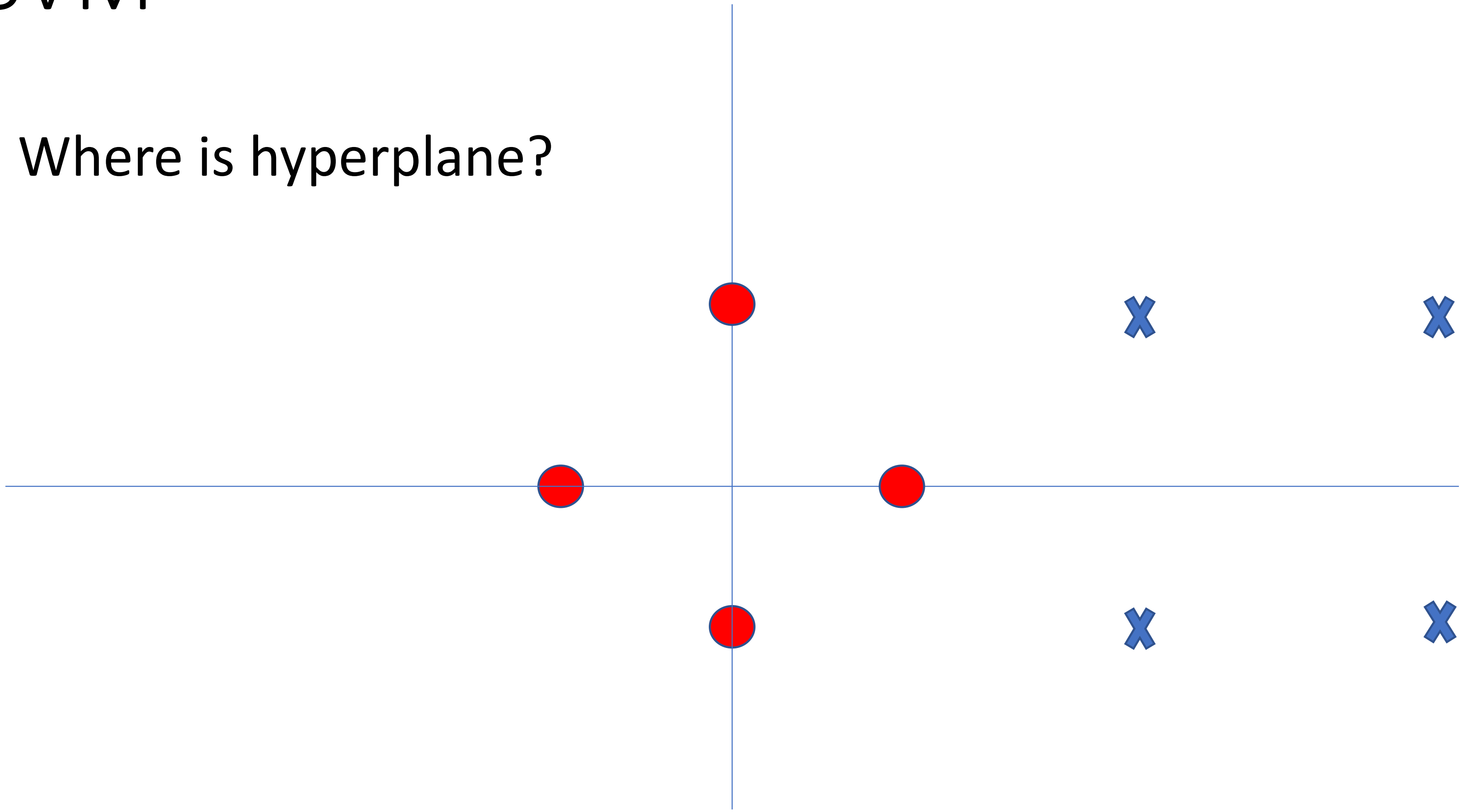
- But how do we get $\vec{y}$?

SVM



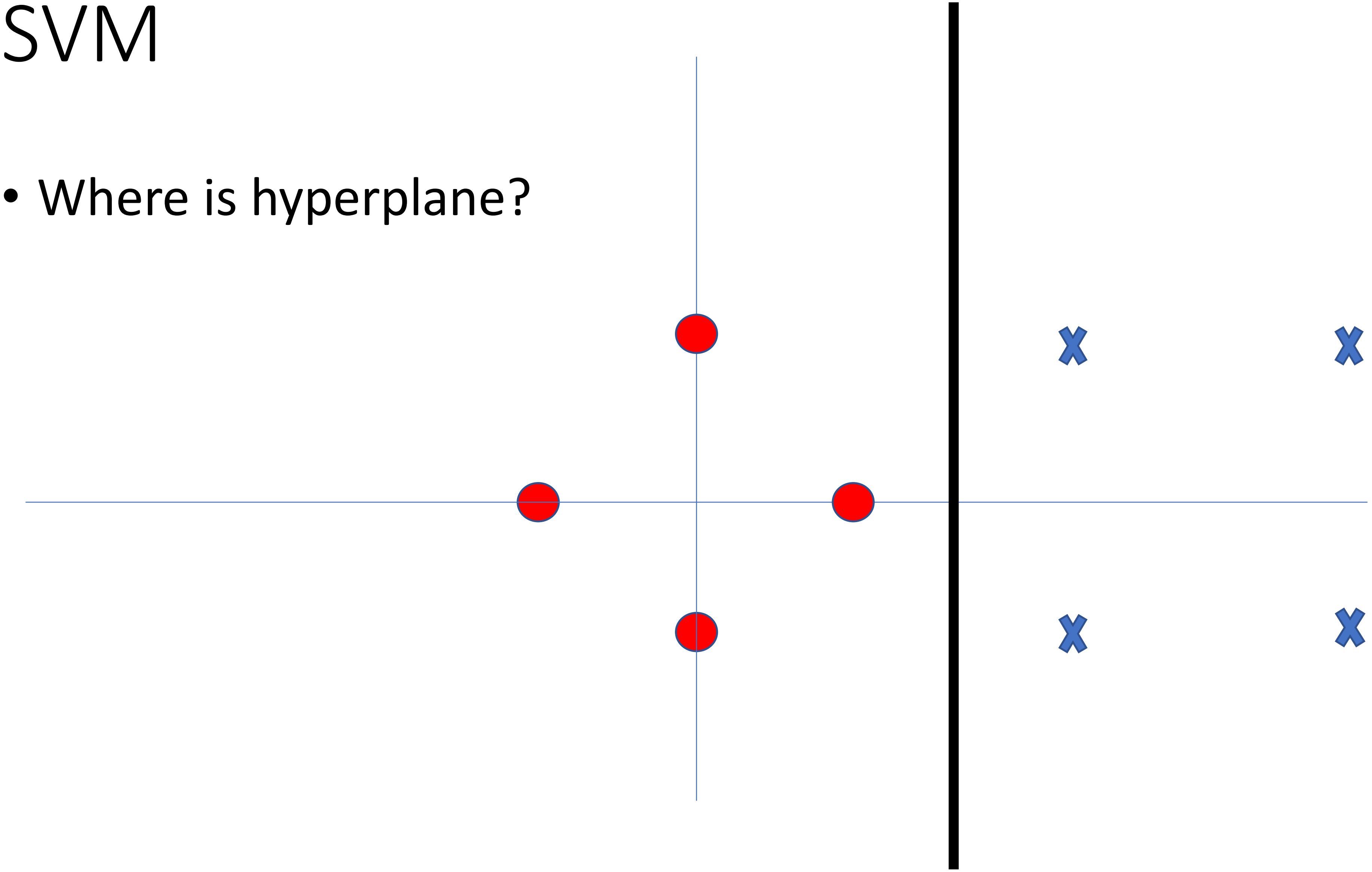
SVM

- Where is hyperplane?



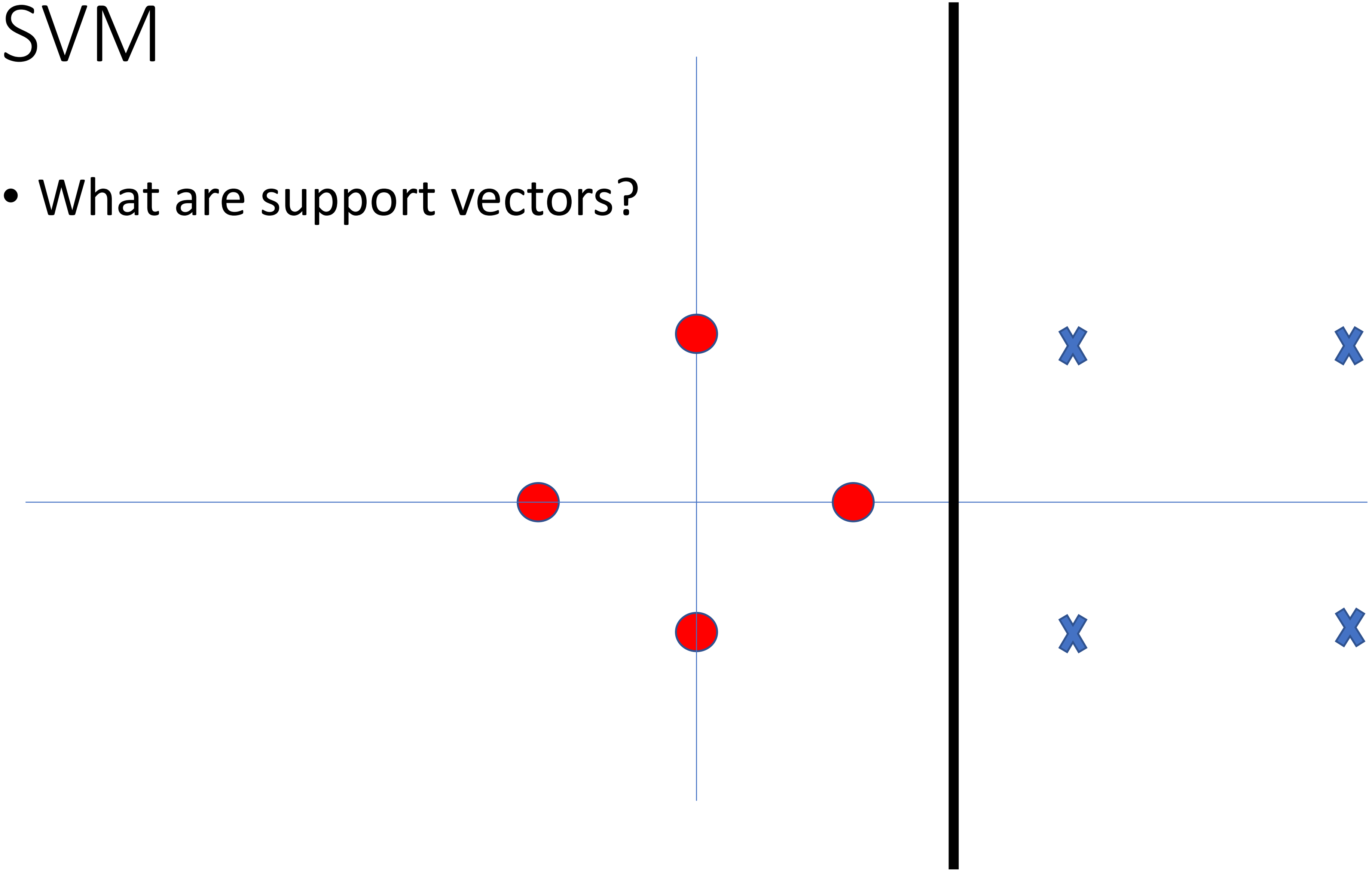
SVM

- Where is hyperplane?



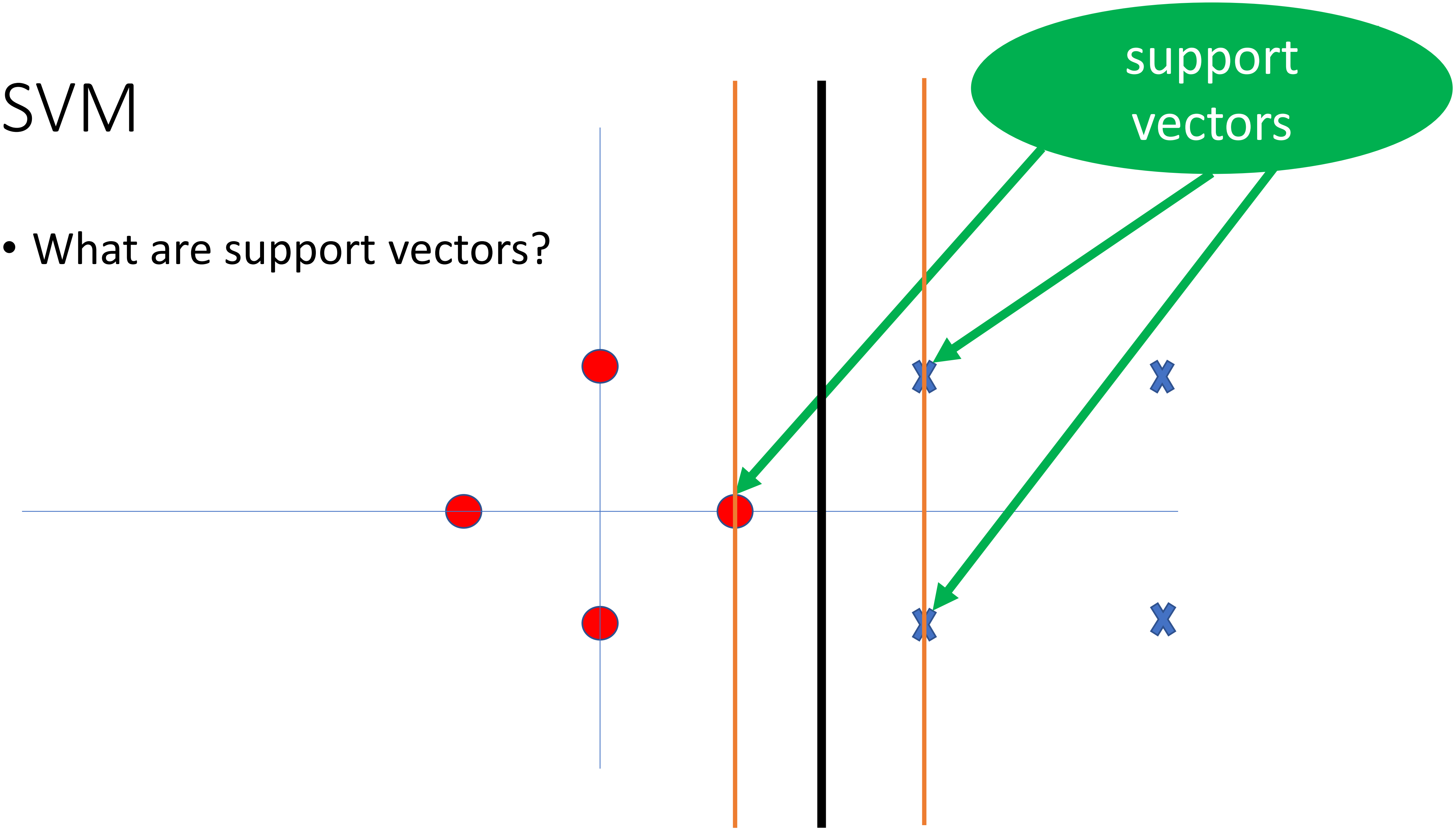
SVM

- What are support vectors?



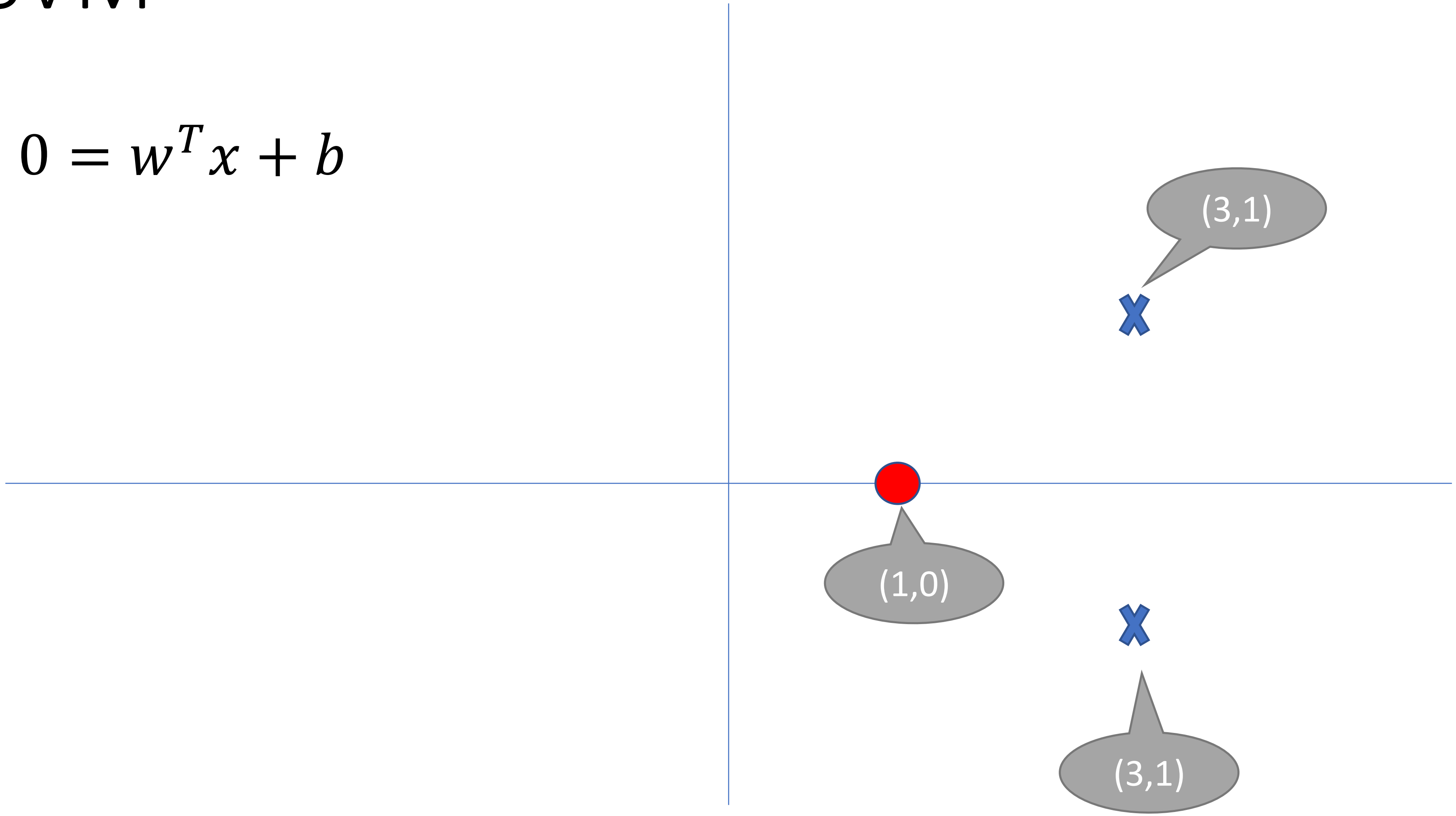
SVM

- What are support vectors?



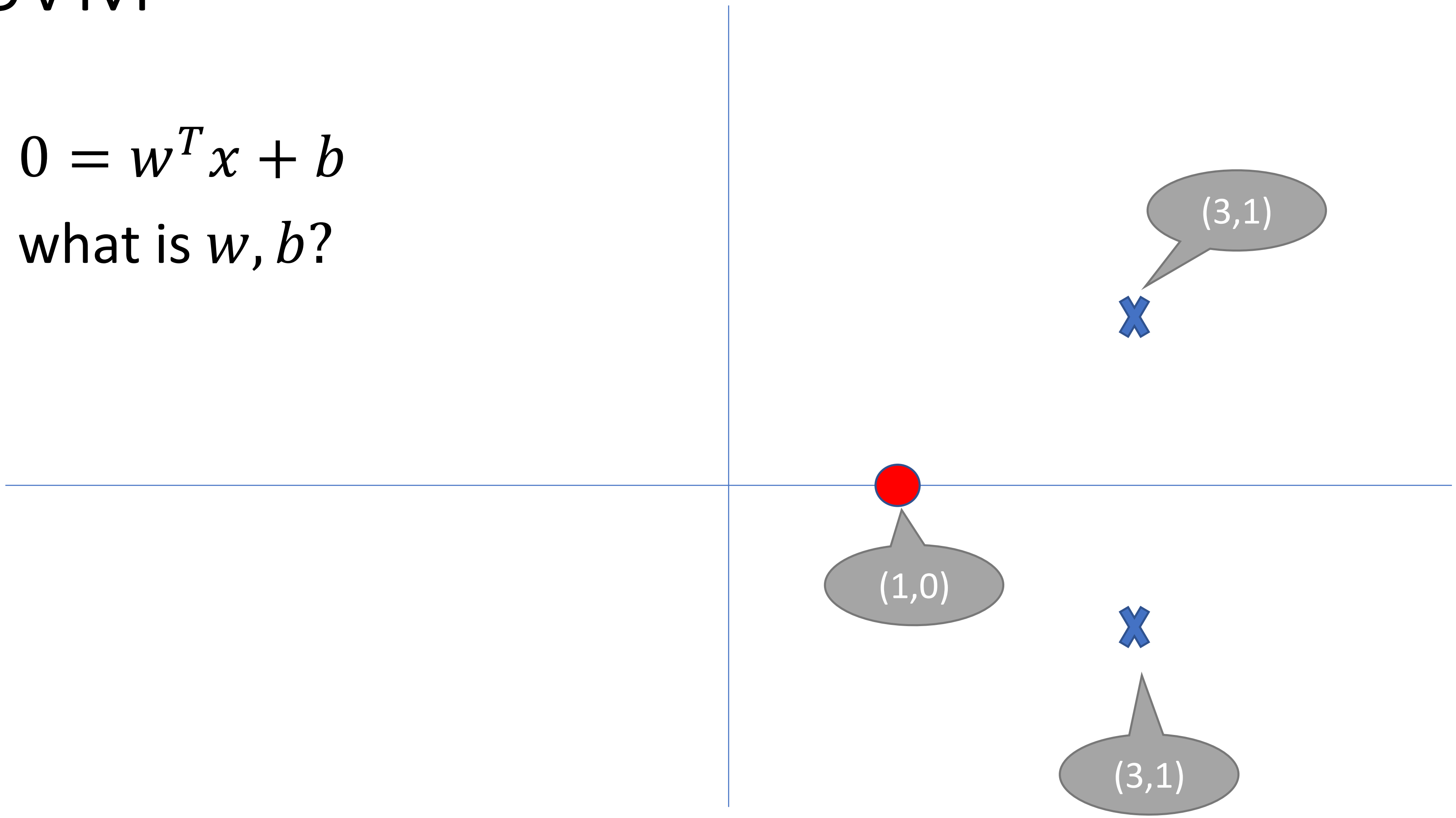
SVM

- $0 = w^T x + b$



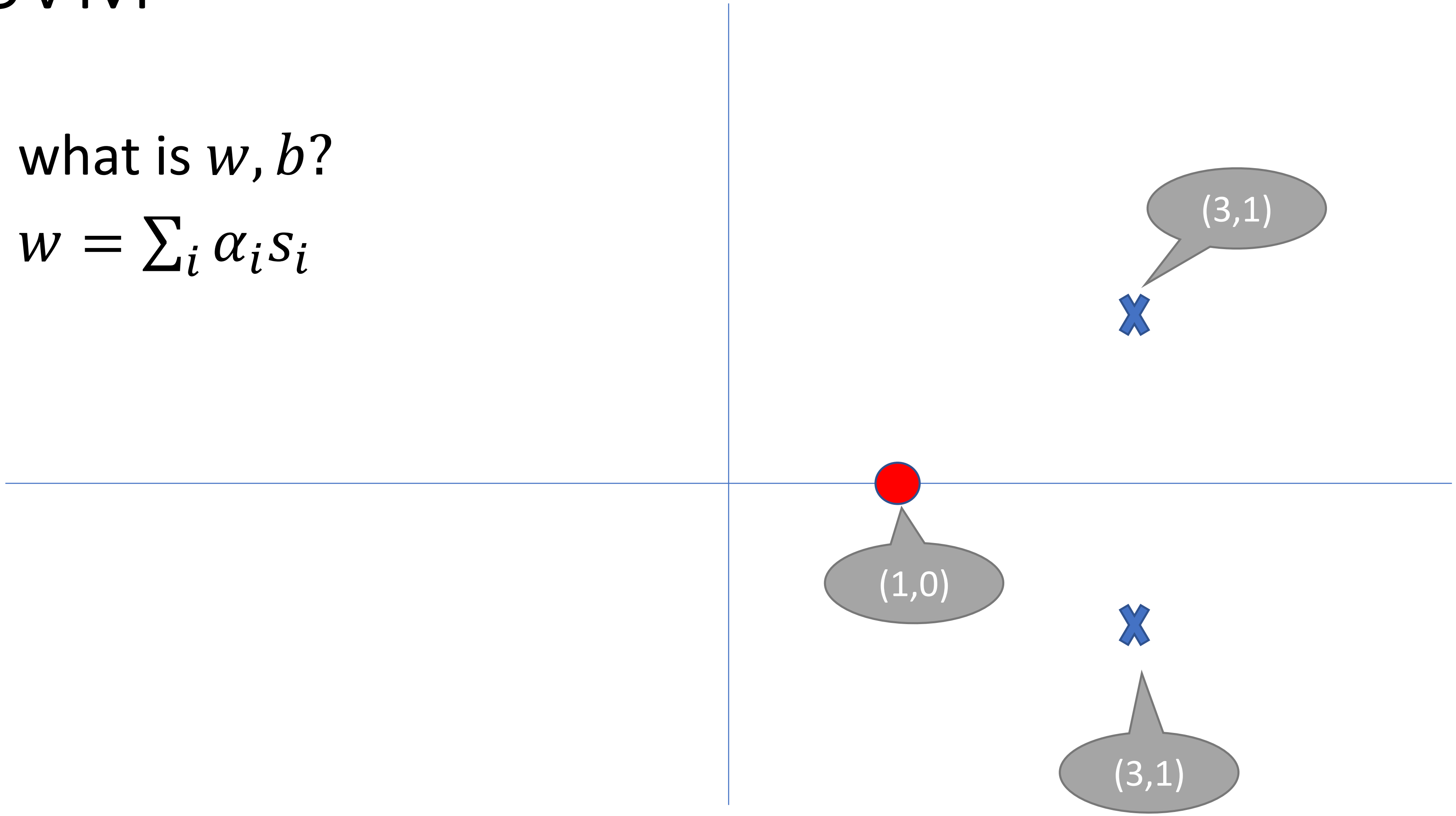
SVM

- $0 = w^T x + b$
- what is w, b?



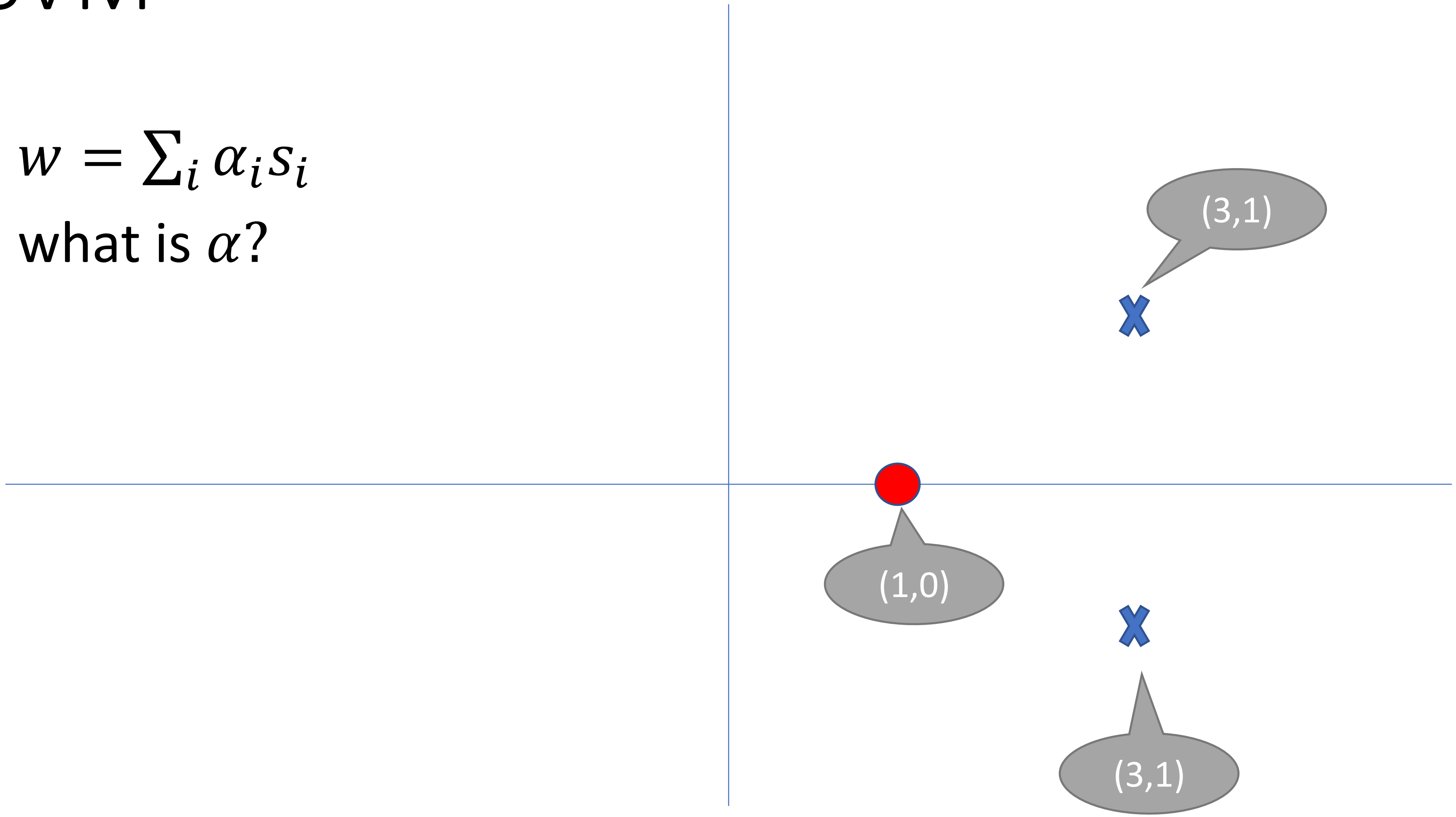
SVM

- what is w, b?
- $w = \sum_i \alpha_i s_i$



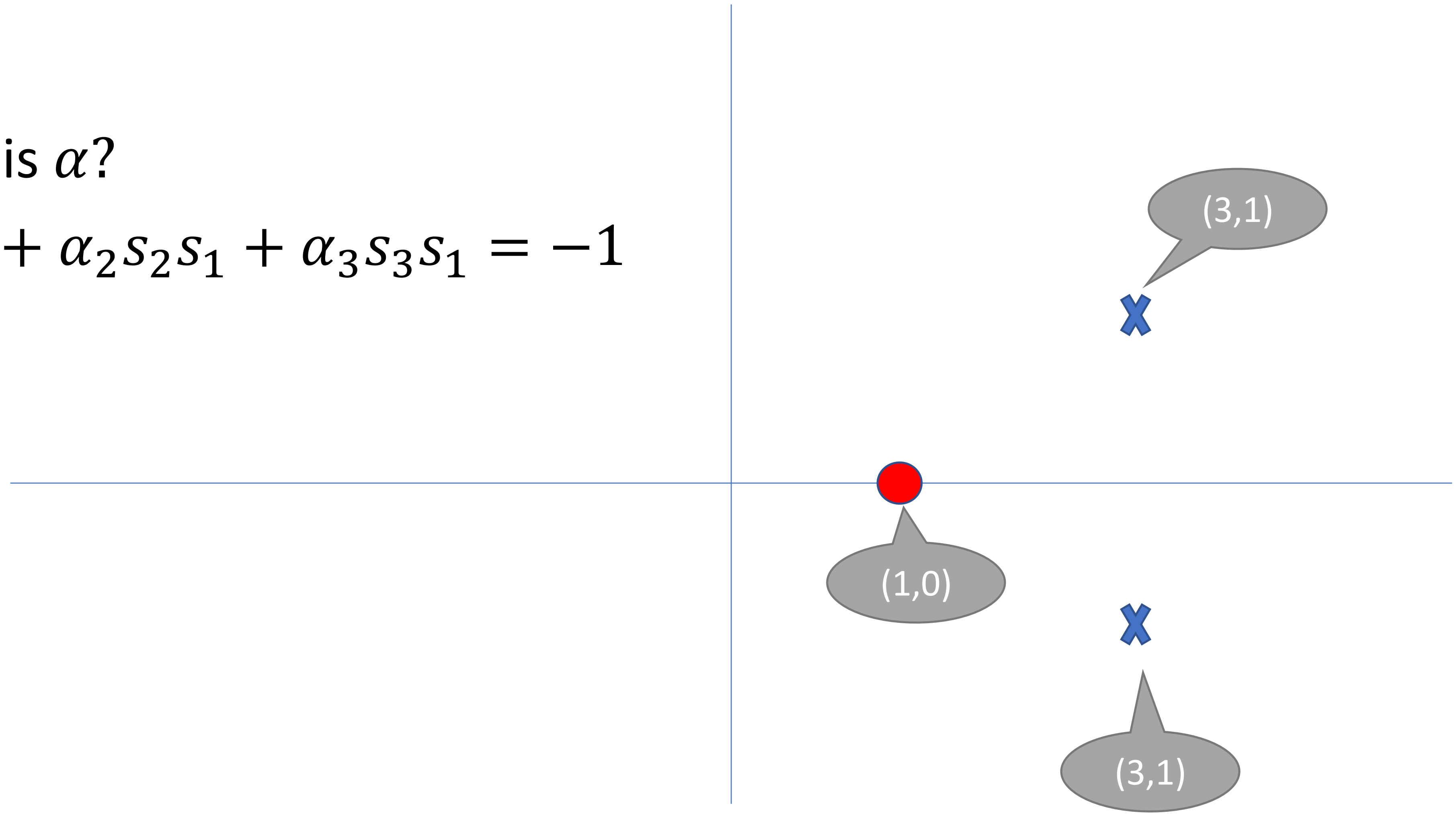
SVM

- $w = \sum_i \alpha_i s_i$
- what is α?



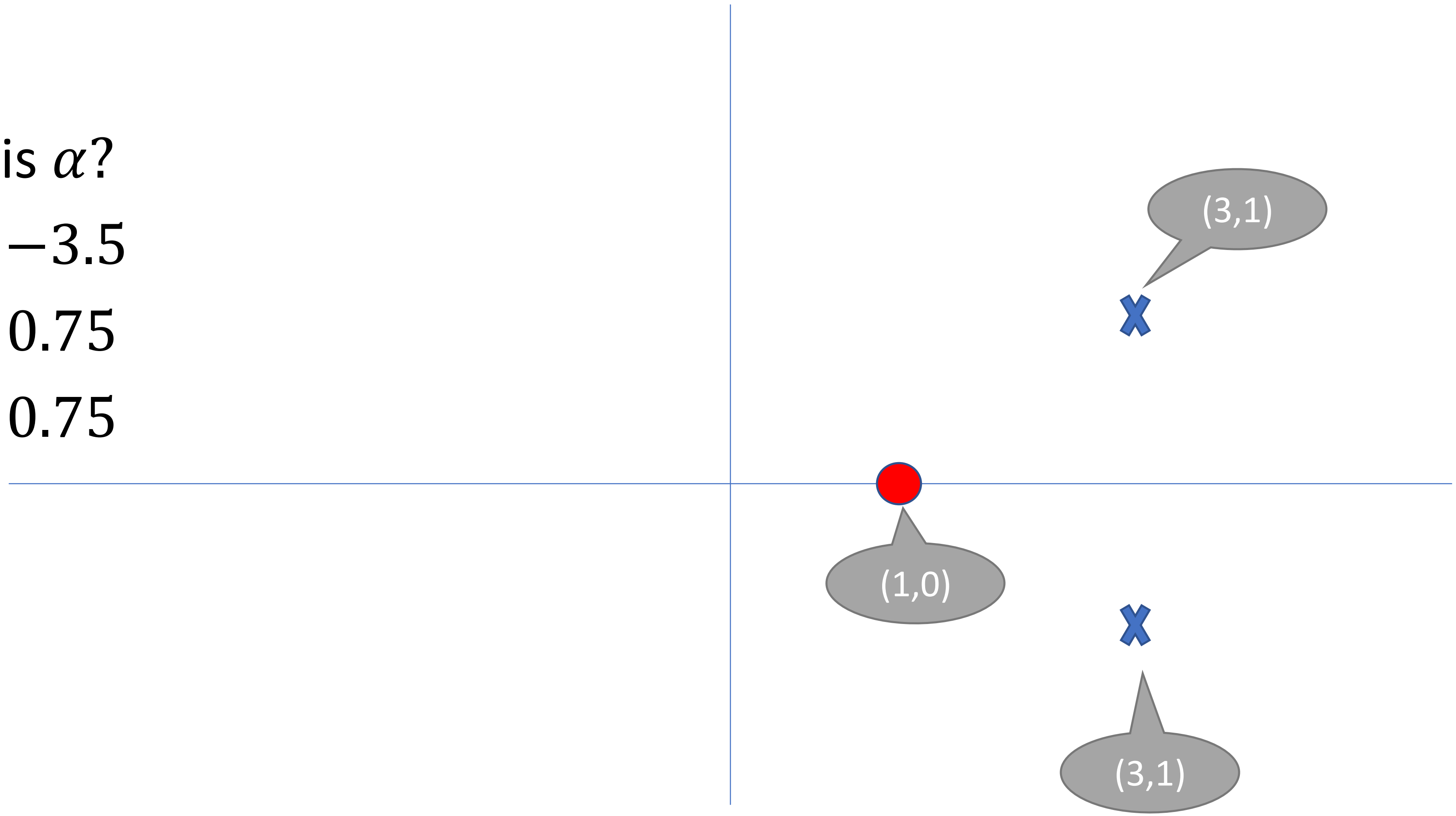
SVM

- what is α?
- $\alpha_1 s_1^2 + \alpha_2 s_2 s_1 + \alpha_3 s_3 s_1 = -1$



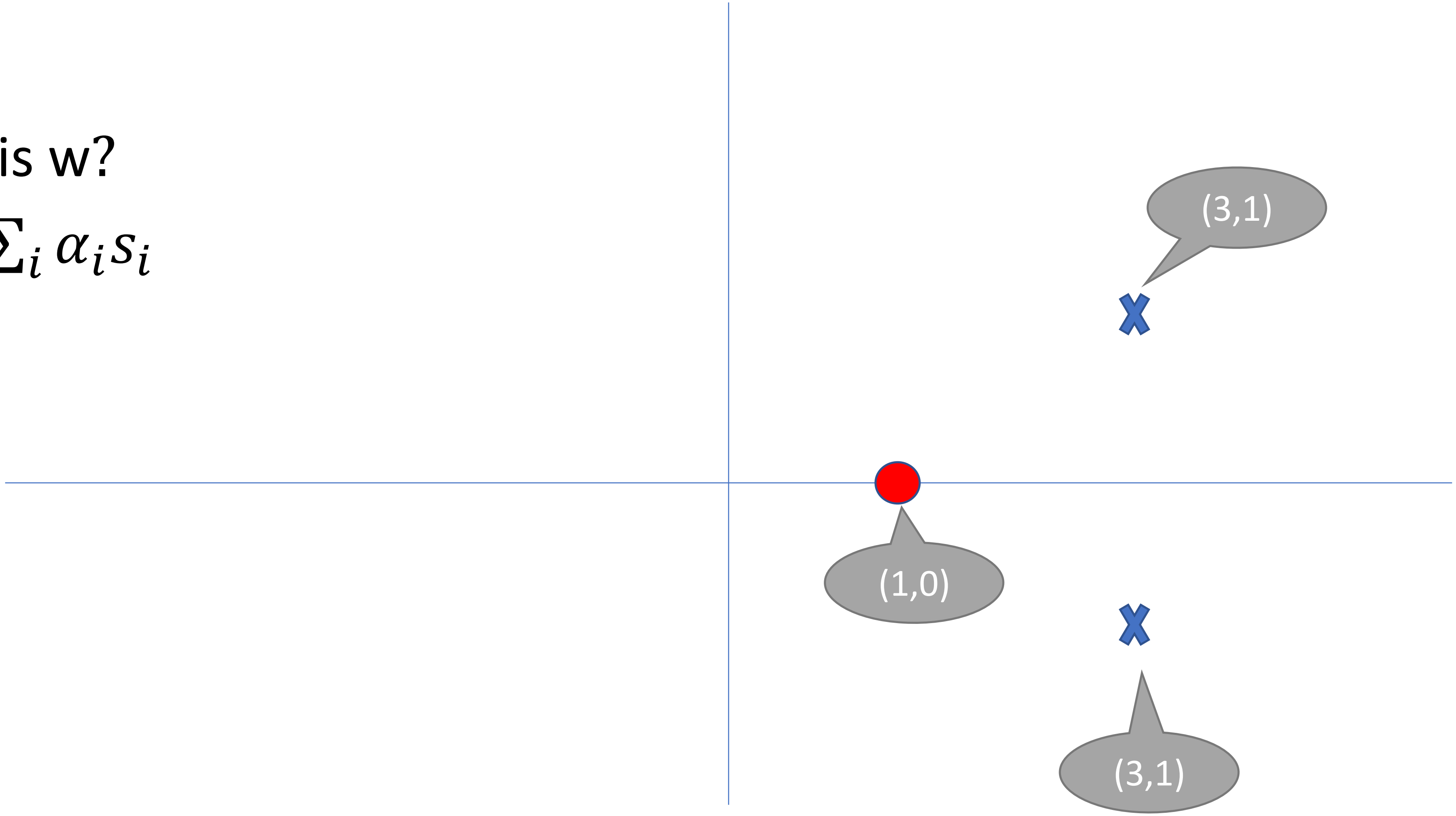
SVM

- what is α?
- $\alpha_1 = -3.5$
- $\alpha_2 = 0.75$
- $\alpha_3 = 0.75$



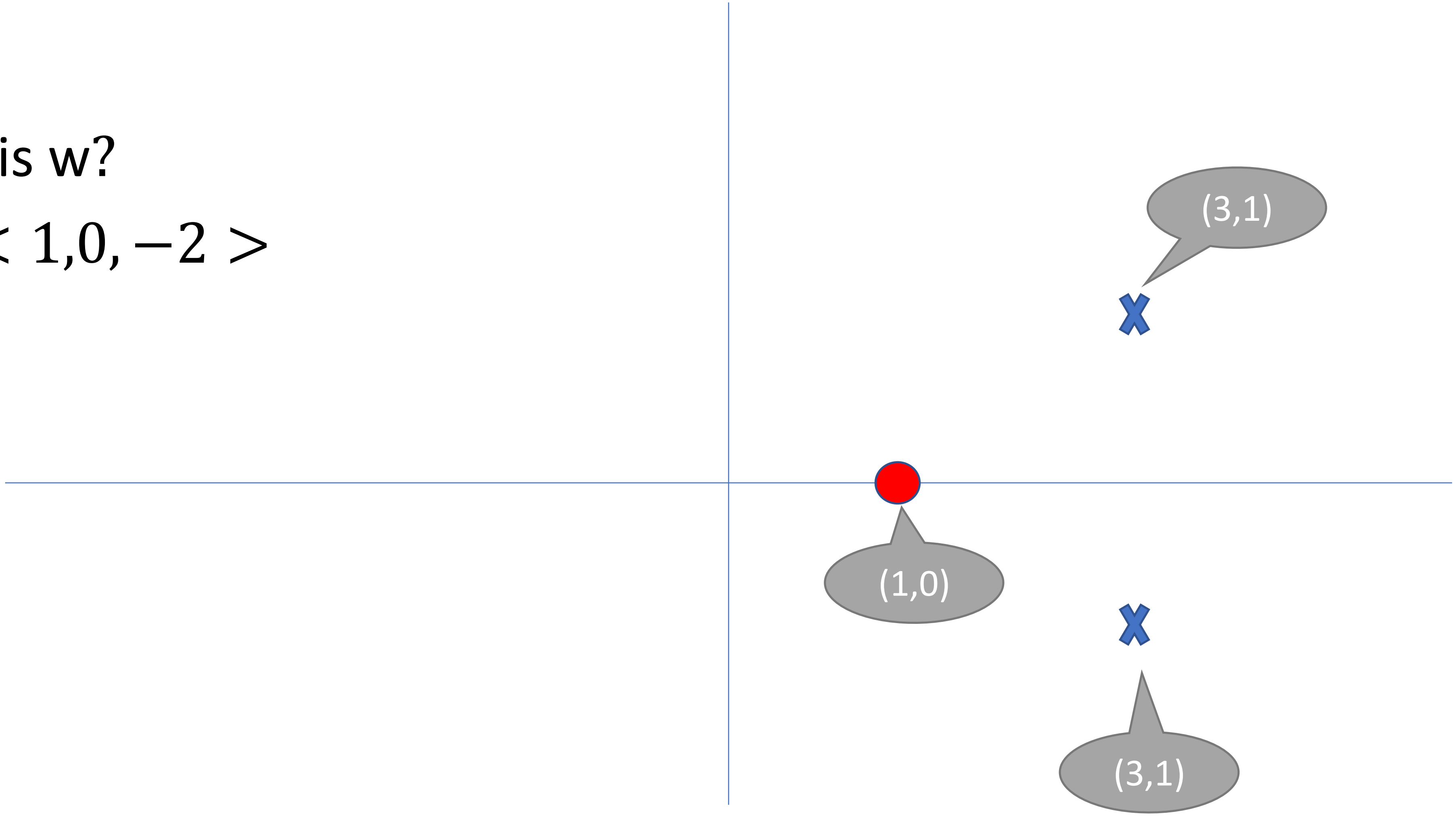
SVM

- what is w?
- $w = \sum_i \alpha_i s_i$



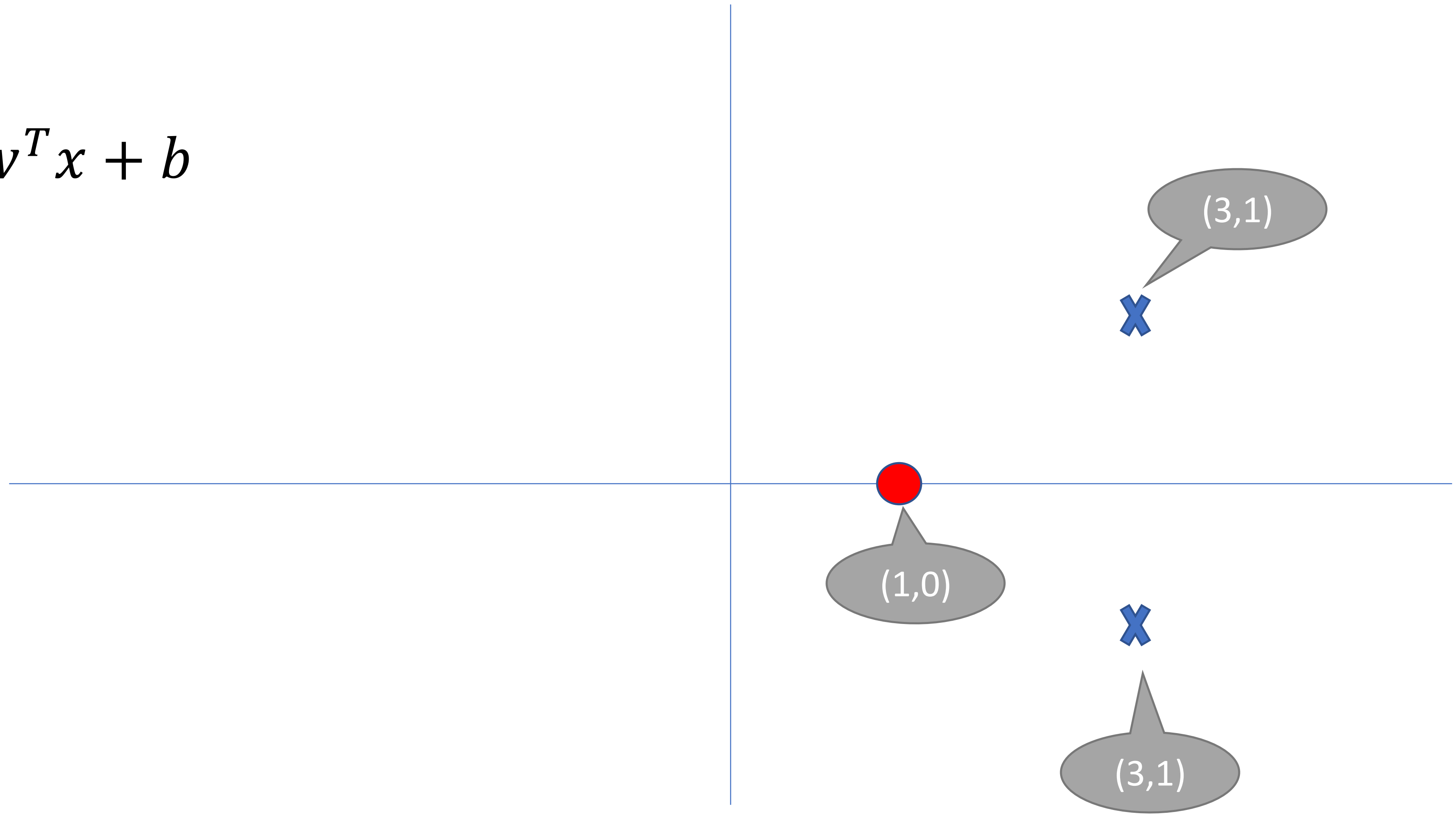
SVM

- what is w?
- $w =< 1,0,-2 >$



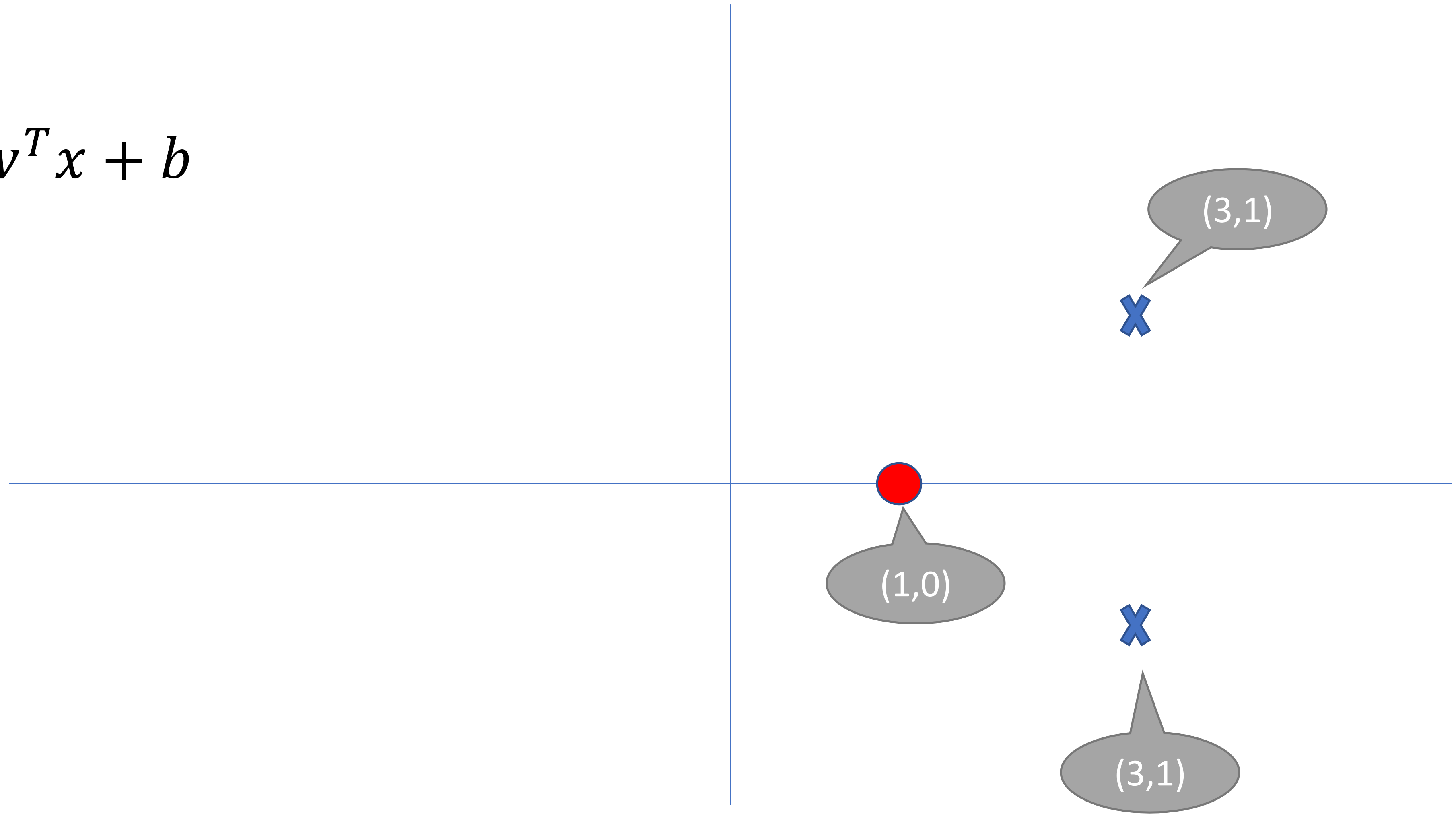
SVM

- $0 = w^T x + b$



SVM

- $0 = w^T x + b$
- x = 2



SVM

- $0 = w^T x + b$
- x = 2