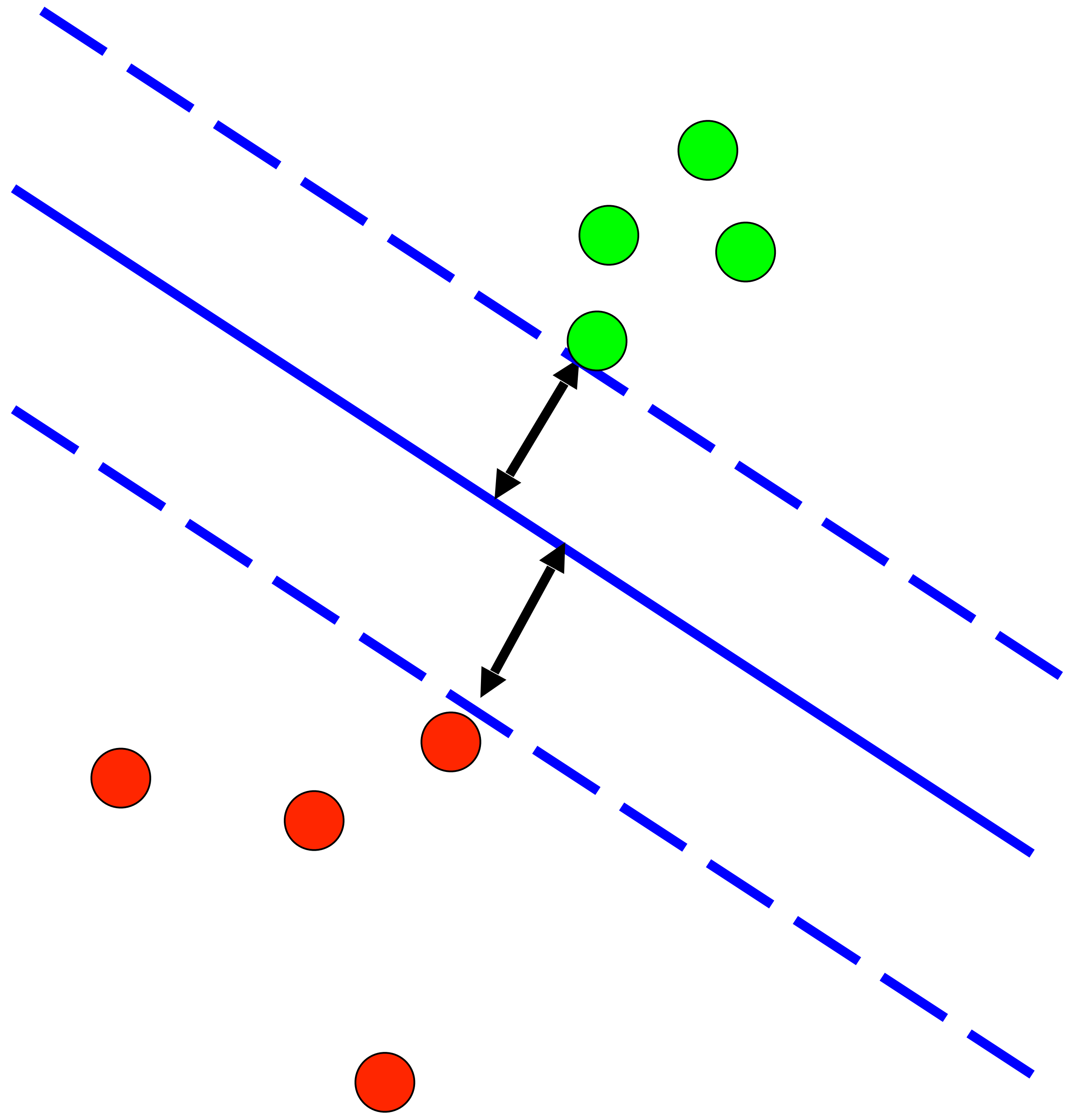
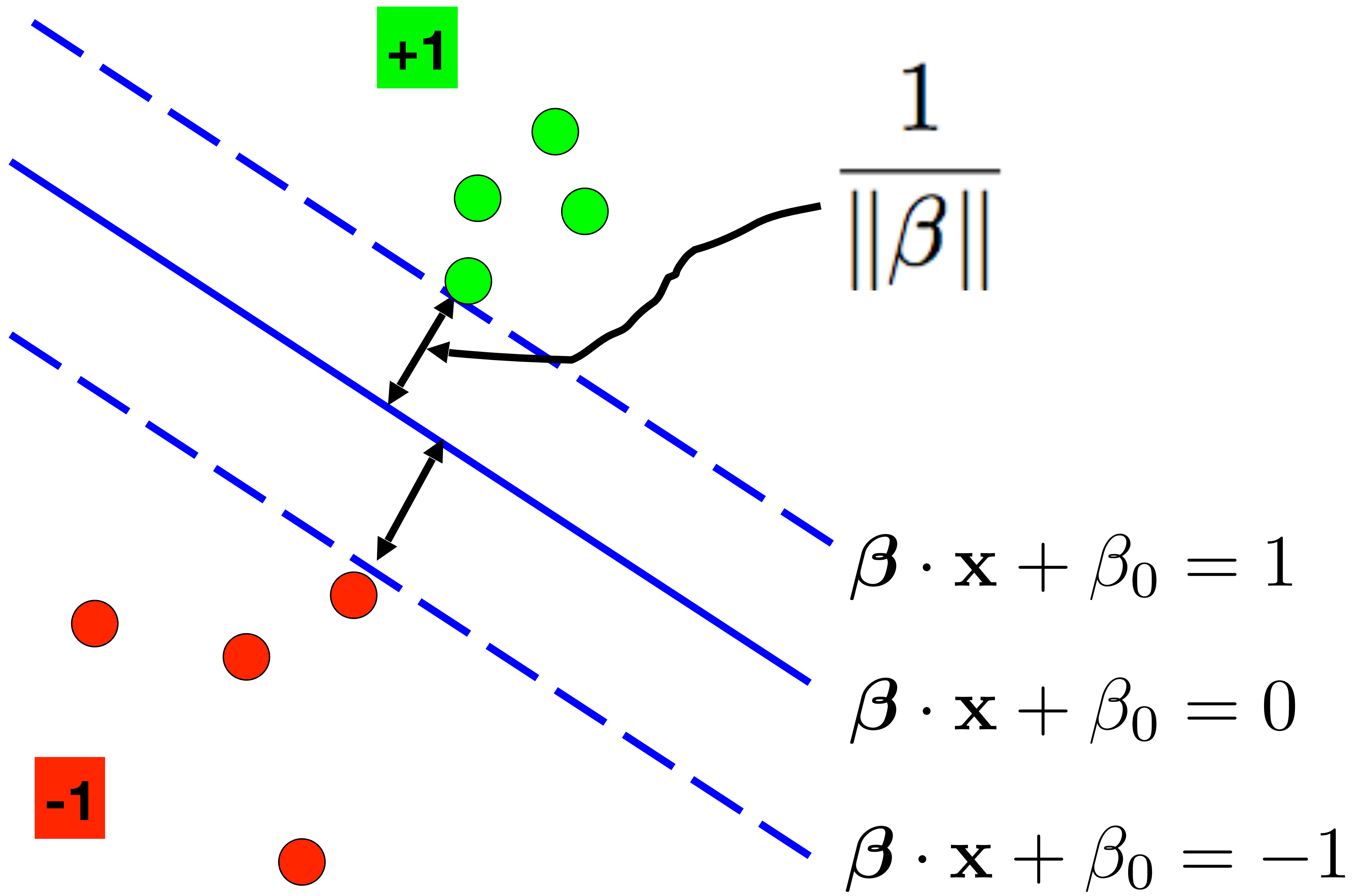


$$\boldsymbol{\beta} \cdot \mathbf{x} + \beta_0 = 0$$

It is desirable to have the width (called margin) between the two lines to be large.

How to formulate this problem?

**Solid Blue Line**: The coefficients (b, b0) are not uniquely determined. We can scale them by any number (pos/neg), the line stays the same.




**It is desirable to have the width (called margin) between the two lines to be large.**

**How to formulate this problem?**

**Solid Blue Line: The coefficients (b, b0) are not uniquely determined. We can multiple them by any number (pos/neg), the line stays the same.**

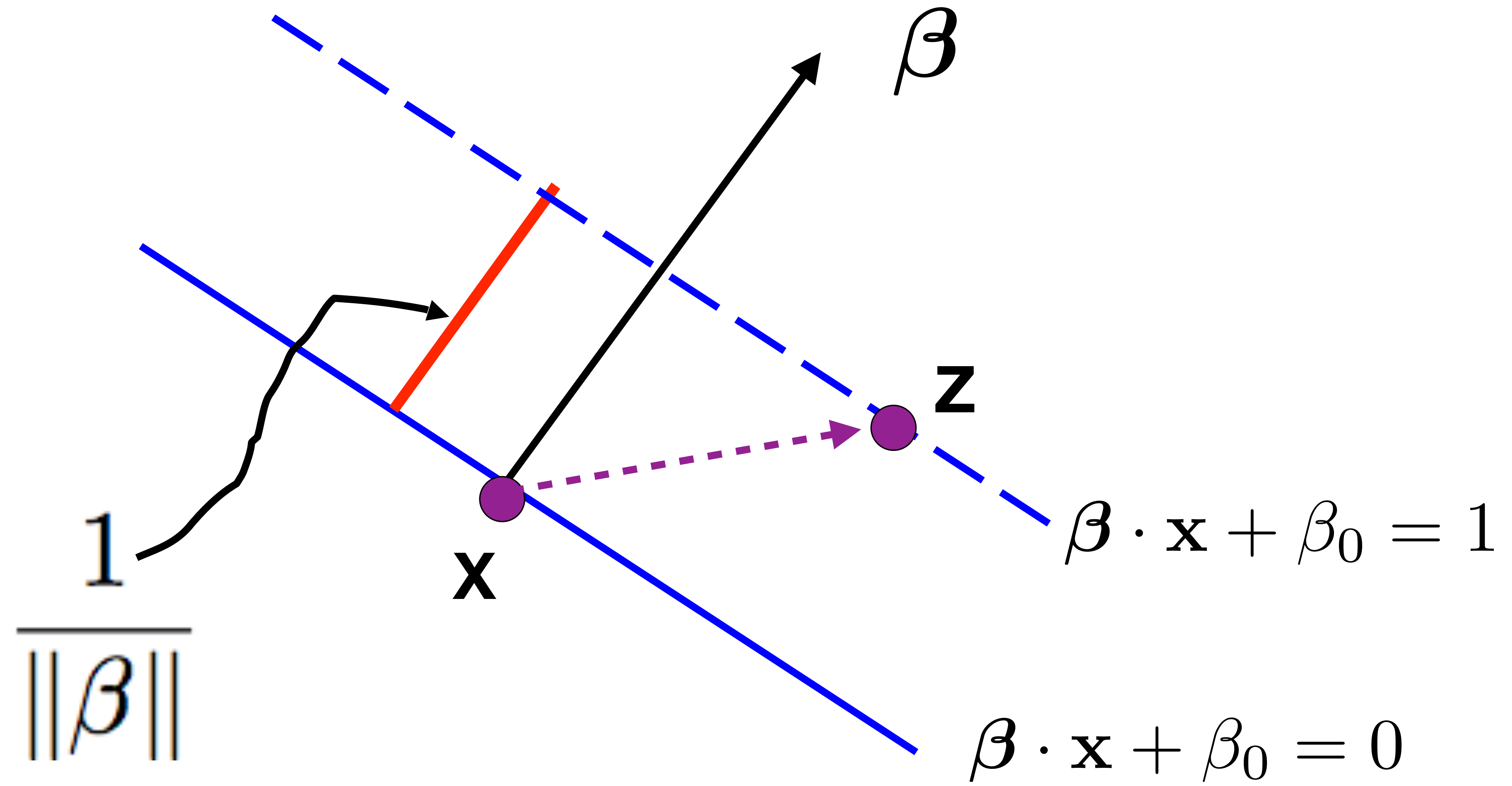
**1. Fix the sign: y = +1 or -1.**

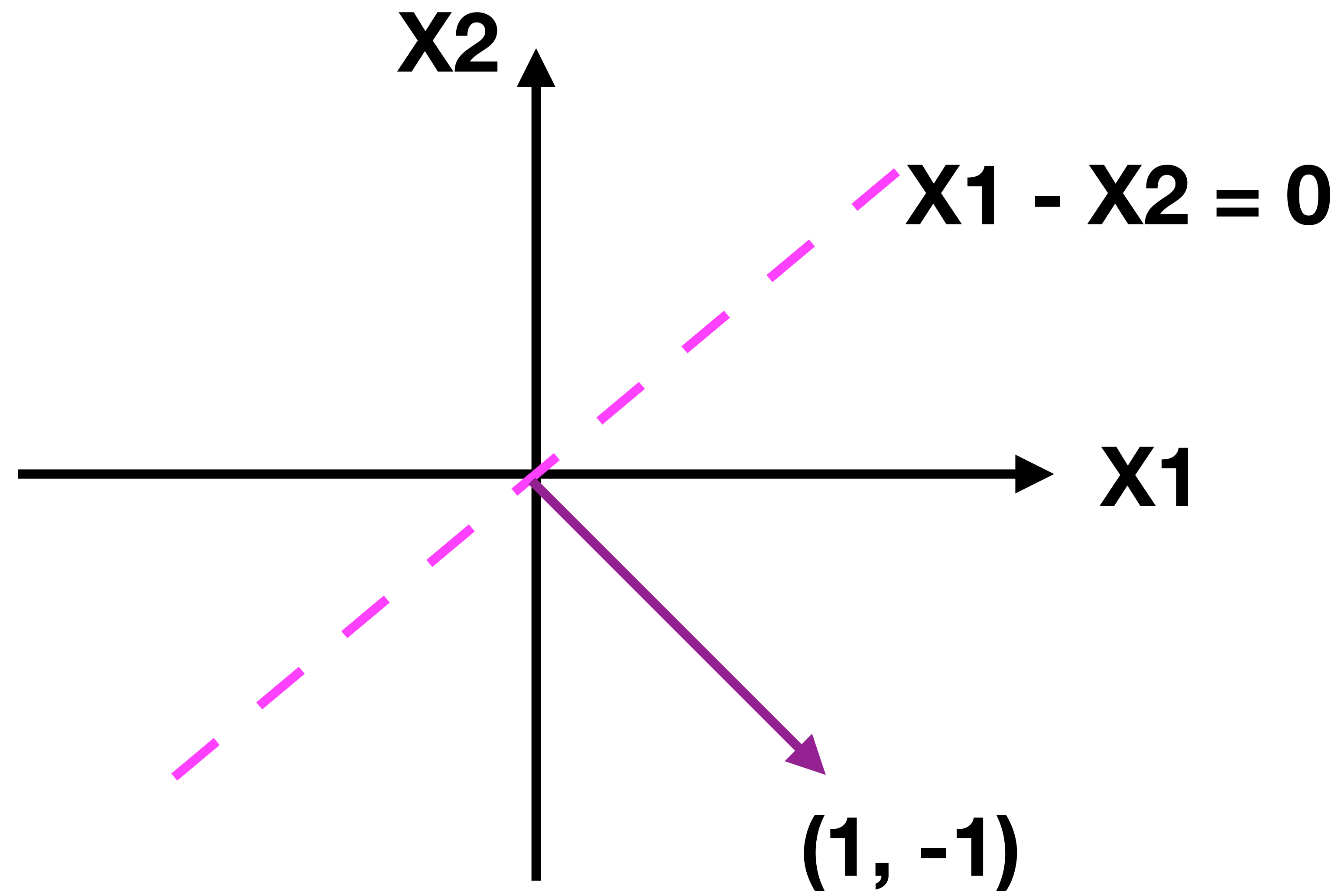
**b*x + b_0 > 0, if y = + 1**
**b*x + b_0 < 0, if y = - 1**

**2. Fix the magnitude: parameterize the two dashed lines as**

**b*x + b_0 = +1**
**b*x + b_0 = -1**

**Two dashed lines determine this wide avenue, and the solid line is in the middle.**



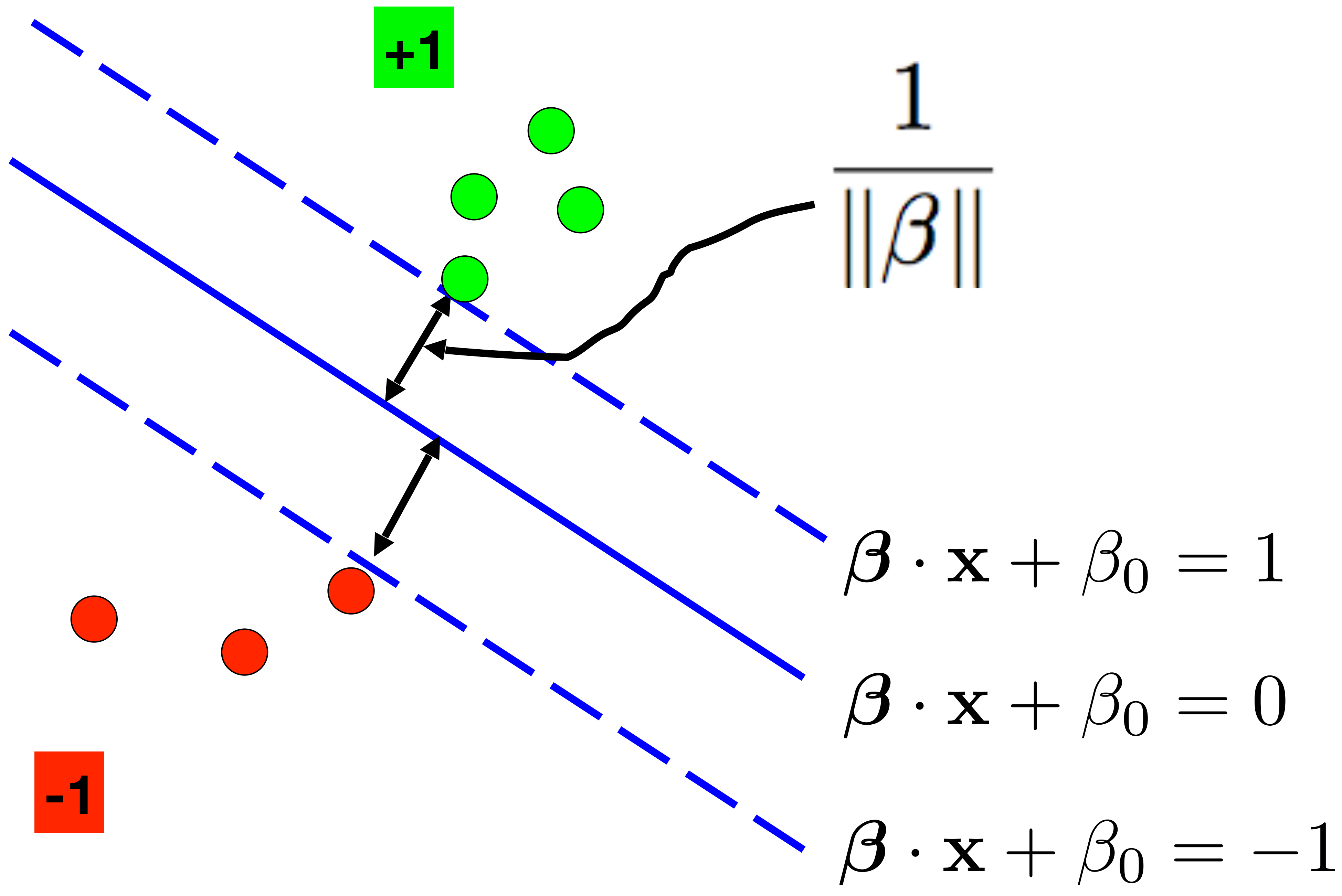




**Line: b*x + b0 =0**
**Interpretation of b: direction that is orthogonal to the line**

**How to compute the distance between these two parallel lines?**

$$(\mathbf{x}-\mathbf{z})^t \frac{\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} = \frac{\mathbf{x}^t\boldsymbol{\beta} - \mathbf{z}^t\boldsymbol{\beta}}{\|\boldsymbol{\beta}\|} = \frac{1}{\|\boldsymbol{\beta}\|}$$

In my calculation, the signs may not be right, but all we care is the magnitude (i.e., we should add absolute value on each expression).

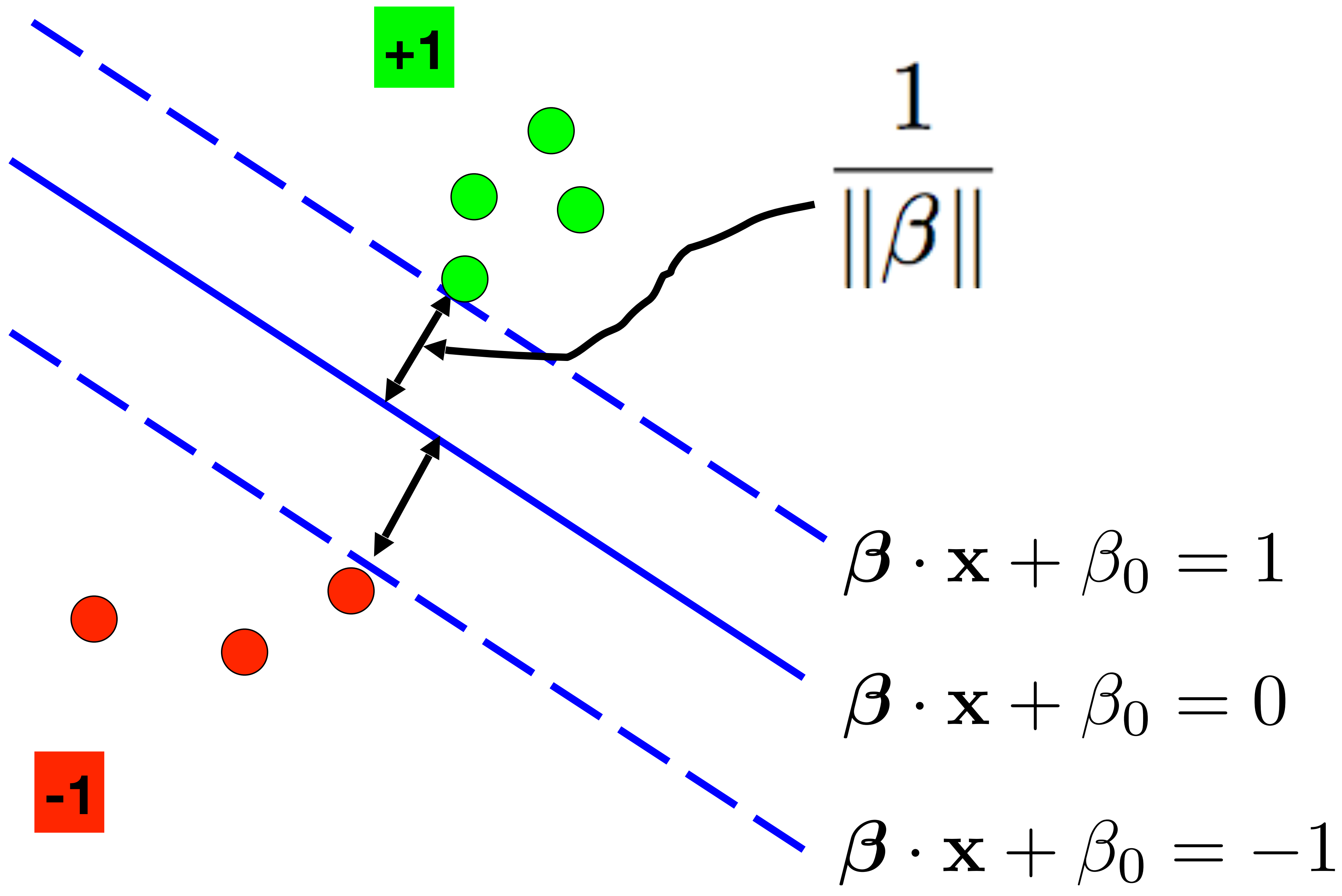



## Max-Margin Problem

$$\min_{\boldsymbol{\beta},\beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 \tag{1}$$
$$\text{subject to} \quad y_i(\boldsymbol{\beta} \cdot \mathbf{x}_i + \beta_0) - 1 \geq 0,$$

where $\boldsymbol{\beta} \cdot \mathbf{x}_i = \boldsymbol{\beta}^t \mathbf{x}_i$ denotes the (Euclidian) inner product between two vectors. The constraints are imposed to make sure that the points are on the correct side of the dashed lines, i.e.,

$$\boldsymbol{\beta} \cdot \mathbf{x}_i + \beta_0 \geq +1 \quad \text{for } y_i = +1,$$
$$\boldsymbol{\beta} \cdot \mathbf{x}_i + \beta_0 \leq -1 \quad \text{for } y_i = -1.$$




- Convex quadratic optimization problem with affine constraints.
- Any local optimum is a global optimum.
- **KKT conditions** are sufficient and necessary
- Equivalence between **the Primal** and **the Dual**.

## Max-Margin Problem

$$\min_{\boldsymbol{\beta},\beta_0} \quad \frac{1}{2}\|\boldsymbol{\beta}\|^2 \tag{1}$$

$$\text{subject to} \quad y_i(\boldsymbol{\beta} \cdot \mathbf{x}_i + \beta_0) - 1 \geq 0,$$

where $\boldsymbol{\beta} \cdot \mathbf{x}_i = \boldsymbol{\beta}^t \mathbf{x}_i$ denotes the (Euclidian) inner product between two vectors. The constraints are imposed to make sure that the points are on the correct side of the dashed lines, i.e.,

$$\boldsymbol{\beta} \cdot \mathbf{x}_i + \beta_0 \geq +1 \quad \text{for } y_i = +1,$$
$$\boldsymbol{\beta} \cdot \mathbf{x}_i + \beta_0 \leq -1 \quad \text{for } y_i = -1.$$

**Primal**

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2$$

$$\text{subj to } y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0,$$
$$i = 1, \ldots, n$$

**Dual**

$$\max_{\lambda_{1:n}} \sum \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subj to } \sum \lambda_i y_i = 0,$$
$$\lambda_i \geq 0$$

**KKT conditions**

$$\sum_i \lambda_i y_i \mathbf{x}_i = \boldsymbol{\beta}$$

$$\sum_i \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0$$

$$\lambda_i \Big[ y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \Big] = 0$$

$$\min_x f(x)$$

**First-order necessary condition**

$$-\frac{\partial f(x)}{\partial x} = \mathbf{0}$$

$$\min_x f(x)$$

**First-order necessary condition**

$$-\frac{\partial f(x)}{\partial x} = \mathbf{0}$$

$$\min_x f(x)$$
$$\text{subj to } g(x) = b$$

$$-\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x}$$

**direction that can reduce f(x)**

**forbidden direction that would violate g(x)=b**

$$\min_x f(x)$$

**First-order necessary condition**

$$-\frac{\partial f(x)}{\partial x} = \mathbf{0}$$

$$\min_x f(x)$$
$$\text{subj to } g(x) = b$$

$$-\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x}$$

**direction that can reduce f(x)**

**forbidden direction that would violate g(x)=b**

$$\min_x f(x)$$
$$\text{subj to } g(x) \geq b$$

$$-\frac{\partial f(x)}{\partial x} = -\lambda \frac{\partial g(x)}{\partial x}$$
$$\lambda \geq 0$$
$$g(x) - b \geq 0$$
$$\lambda(g(x) - b) = 0$$

If x is a local optimum for the constrained optimization, then it must satisfy the **KKT conditions**.

- x is active (lambda >= 0)
- x is inactive (lambda = 0)

$$\min_x f(x)$$

**First-order necessary condition**

$$-\frac{\partial f(x)}{\partial x} = \mathbf{0}$$

$$\min_x f(x) \quad \text{subj to } g(x) = b$$

$$-\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial g(x)}{\partial x}$$

$$\min_x f(x) \quad \text{subj to } g(x) \geq b$$

$$-\frac{\partial f(x)}{\partial x} = -\lambda \frac{\partial g(x)}{\partial x}$$

$$\lambda \geq 0$$

$$g(x) - b \geq 0$$

$$\lambda(g(x) - b) = 0$$

$$\text{Define } L(x, \lambda) = f(x) - \lambda\big(g(x) - b\big)$$

$$\boxed{\frac{\partial}{\partial x} L = 0}$$

If x is a local optimum for the constrained optimization, then it must satisfy the **KKT conditions**.

- x is active (lambda >= 0)
- x is inactive (lambda = 0)

### Primal

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2$$

$$\text{subj to } y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0,$$

$$i = 1, \ldots, n$$

### Dual

$$\max_{\lambda_{1:n}} \sum \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subj to } \sum \lambda_i y_i = 0,$$

$$\lambda_i \geq 0$$

$$\frac{\partial L}{\partial x} = 0$$

$$\lambda \geq 0$$

$$g(x) \geq b$$

$$\lambda(g(x) - b) = 0$$

### KKT conditions

$$\sum_i \lambda_i y_i \mathbf{x}_i = \boldsymbol{\beta}$$

$$\sum_i \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0$$

$$\lambda_i \left[ y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \right] = 0$$

### Lagrange function

$$L(\boldsymbol{\beta}, \beta_0, \lambda_{1:n})$$

$$= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_i \lambda_i \left[ y_i(\mathbf{x}_i^t \boldsymbol{\beta} + \beta_0) - 1 \right]$$

$$= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_i \lambda_i y_i(\mathbf{x}_i^t \boldsymbol{\beta} + \beta_0) + \sum_i \lambda_i$$

**Primal**

$$\boxed{\begin{array}{l} \min_x f(x) \\ \text{subj to } g(x) \geq b \end{array}}$$

$$\updownarrow$$

$$\boxed{\min_x \max_{\lambda \geq 0} \Big[ f(x) - \lambda\big(g(x) - b\big) \Big]}$$

$L(x, \lambda)$

$$\max_{\lambda \geq 0} \Big[ f(x) - \lambda\big(g(x) - b\big) \Big] = \begin{cases} f(x) & \text{if } g(x) \geq b \\ \infty & \text{if } g(x) < b \end{cases}$$

## Primal

$$\boxed{\begin{aligned}&\min_x f(x)\\ &\text{subj to } g(x) \geq b\end{aligned}}$$

$$\boxed{\min_x \max_{\lambda \geq 0} \left[\underbrace{f(x) - \lambda\big(g(x) - b\big)}_{L(x,\lambda)}\right]}$$

$$\max_{\lambda \geq 0} \left[f(x) - \lambda\big(g(x) - b\big)\right] = \begin{cases} f(x) & \text{if } g(x) \geq b \\ \infty & \text{if } g(x) < b \end{cases}$$

**Under some conditions that are satisfied here, we have**

$$\boxed{\min_x \max_\lambda L(x,\lambda) = \max_\lambda \min_x L(x,\lambda) = L(x^*,\lambda^*)}$$

**Primal**

$$\min_x f(x)$$
$$\text{subj to } g(x) \geq b$$

$$\min_x \max_{\lambda \geq 0} \left[ f(x) - \lambda\big(g(x) - b\big) \right]$$

$L(x, \lambda)$

**Dual**

$$\max_{\lambda \geq 0} \min_x \left[ f(x) - \lambda\big(g(x) - b\big) \right]$$

**Equivalent and KKT conditions can link the two sets of solutions: x* and lambda***

$$\max_{\lambda \geq 0} \left[ f(x) - \lambda\big(g(x) - b\big) \right] = \begin{cases} f(x) & \text{if } g(x) \geq b \\ \infty & \text{if } g(x) < b \end{cases}$$

**Under some conditions that are satisfied here, we have**

$$\min_x \max_\lambda L(x, \lambda) = \max_\lambda \min_x L(x, \lambda) = L(x^*, \lambda^*)$$

## Primal

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2$$

$$\text{subj to } y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0,$$
$$i = 1, \ldots, n$$

## Dual

$$\max_{\lambda_{1:n}} \sum \lambda_i - \frac{1}{2}\sum_{i,j} \lambda_i \lambda_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{subj to } \sum \lambda_i y_i = 0,$$
$$\lambda_i \geq 0$$

## KKT conditions

$$\sum_i \lambda_i y_i \mathbf{x}_i = \boldsymbol{\beta}$$

$$\sum_i \lambda_i y_i = 0$$

$$\lambda_i \geq 0$$

$$y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0$$

$$\lambda_i \left[ y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \right] = 0$$

## Lagrange function

$$L(\boldsymbol{\beta}, \beta_0, \lambda_{1:n})$$
$$= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_i \lambda_i \left[ y_i(\mathbf{x}_i^t \boldsymbol{\beta} + \beta_0) - 1 \right]$$
$$= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_i \lambda_i y_i(\mathbf{x}_i^t \boldsymbol{\beta} + \beta_0) + \sum_i \lambda_i$$

## Primal

$$\min_{\boldsymbol{\beta},\beta_0} \frac{1}{2}\|\boldsymbol{\beta}\|^2$$
$$\text{subj to } y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0,$$
$$i = 1, \ldots, n$$

## Dual

$$\max_{\lambda_{1:n}} \sum \lambda_i - \frac{1}{2}\sum_{i,j} \lambda_i \lambda_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$
$$\text{subj to } \sum \lambda_i y_i = 0,$$
$$\lambda_i \geq 0$$

## KKT conditions

$$\sum_i \lambda_i y_i \mathbf{x}_i = \boldsymbol{\beta}$$
$$\sum_i \lambda_i y_i = 0$$
$$\lambda_i \geq 0$$
$$y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \geq 0$$
$$\lambda_i \left[ y_i(\mathbf{x}_i \cdot \boldsymbol{\beta} + \beta_0) - 1 \right] = 0$$

## Lagrange function

$$L(\boldsymbol{\beta}, \beta_0, \lambda_{1:n})$$
$$= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_i \lambda_i \left[ y_i(\mathbf{x}_i^t \boldsymbol{\beta} + \beta_0) - 1 \right]$$
$$= \frac{1}{2}\|\boldsymbol{\beta}\|^2 - \sum_i \lambda_i y_i(\mathbf{x}_i^t \boldsymbol{\beta} + \beta_0) + \sum_i \lambda_i$$

**Why work with Dual?**
1. Easier to solve
2. Many lambda_i's are zero
3. Leads to kernel trick