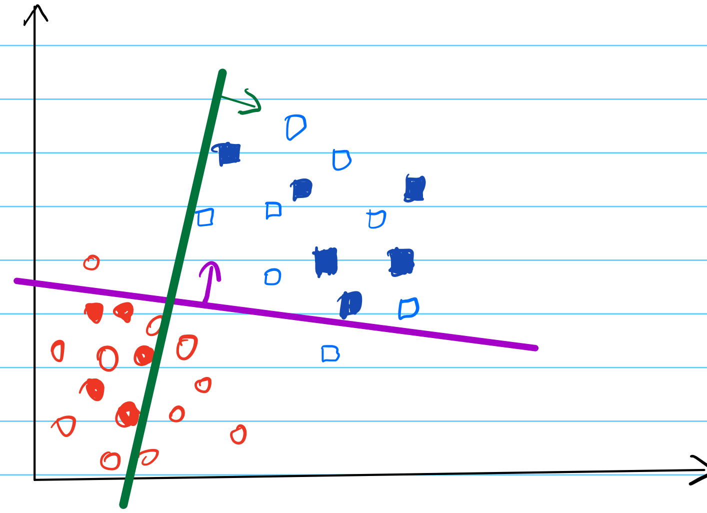
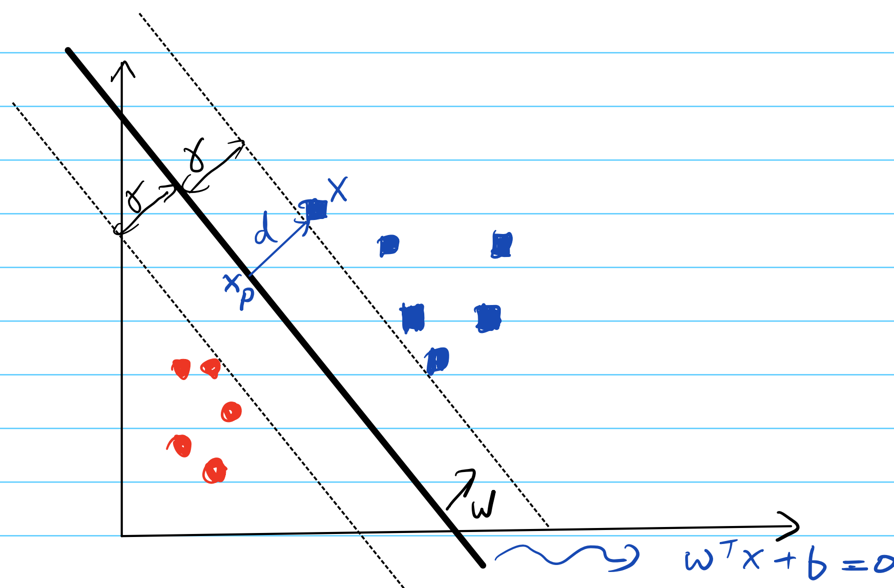
CS 3780/5780
Linear Support Vector Machines



Which Hyperplane do we pick?



Which Hyperplane do we pick?

1. $x_p = x - d$

2. $d = \alpha w$

3. $w^T x_p + b = 0$

4. $\alpha = \frac{w^T x + b}{\|w\|^2}$

5. $\gamma = \|d\| = \alpha \|w\| = \dfrac{|w^T x + b|}{\|w\|^2} \|w\|$

$= \dfrac{|w^T x + b|}{\|w\|}$

$$\gamma(w, b) = \min_{x \in D_{Train}} \frac{|w^T x + b|}{\|w\|}$$

Max Margin Classifier:

$\underset{w,b}{\text{argmax}} \ \gamma(w, b)$ Maximize margin

s.t. $\forall x, y \in D$

$(w^T x + b) y \geq 0$ } get all labels correct

1. $\gamma(\beta w, \beta b) = \gamma(w, b) \quad \forall \beta > 0$?

2. $(w^T x + b) y \geq 0 \Rightarrow \beta (w^T x + b) \geq 0 \quad \forall \beta > 0$

So we can arbitrarily scale $w^T x + b$ by any β

choose scale so that

$$\min_{x \in D} |w^T x + b| = 1$$

Max margin classifier

$$\underset{w,b}{\arg\max} \; \min_{x \in D} \frac{|w^T x + b|}{\|w\|} \quad \text{s.t. } \forall x, y \in D \quad (w^T x + b) y \geq 0$$

$$\Updownarrow$$

$$\underset{w,b}{\arg\max} \; \frac{1}{\|w\|} \quad \text{s.t. } \forall x, y \in D$$

$$y(w^T x + b) \geq 0$$

$$\min_{x \in D} |w^T x + b| = 1$$

$$\Updownarrow$$

$$\underset{w,b}{\arg\min} \; \|w\|^2 \quad \text{st } \forall x, y \in D$$

$$y(w^T x + b) \geq 0$$

$$\min_{x,y \in D} |w^T x + b| = 1$$

??? $\Updownarrow$

$$\underset{w,b}{\arg\min} \; \|w\|^2$$

$$\text{st. } \forall x, y \in D \quad y(w^T x + b) \geq 1$$

Quadratic cost, linear constraints.

Points $x, y \in D$ for which

$|w^T x + b| = 1$ are support vectors.

SVM with soft margin

What if data is not linearly separable?

Allow violations with slack variables ξ_i's

C large what happens ?

$$\arg\min_{w, b, \xi_1 \dots \xi_n} \|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad \forall i \quad y_i (w^T x_i + b) \geq 1 - \xi_i$$

$$\forall i \quad \xi_i \geq 0$$

Alternative formulation:

since we want the smallest $\sum_{i=1}^{n} \xi_i$.

each ξ_i if can be made 0 is 0

else $\xi_i = 1 - y_i (w^T x_i + b)$

Hence $\xi_i = \max\{0, 1 - y_i (w^T x_i + b)\}$

SVM : $\arg\min_{w, b} \|w\|^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i (w^T x_i + b))$

Hinge loss: $\ell(h_{w,b}(x), y) = \max(0, 1 - y\, h_{w,b}(x))$

where $h_{w,b}(x) = w^T x + b$

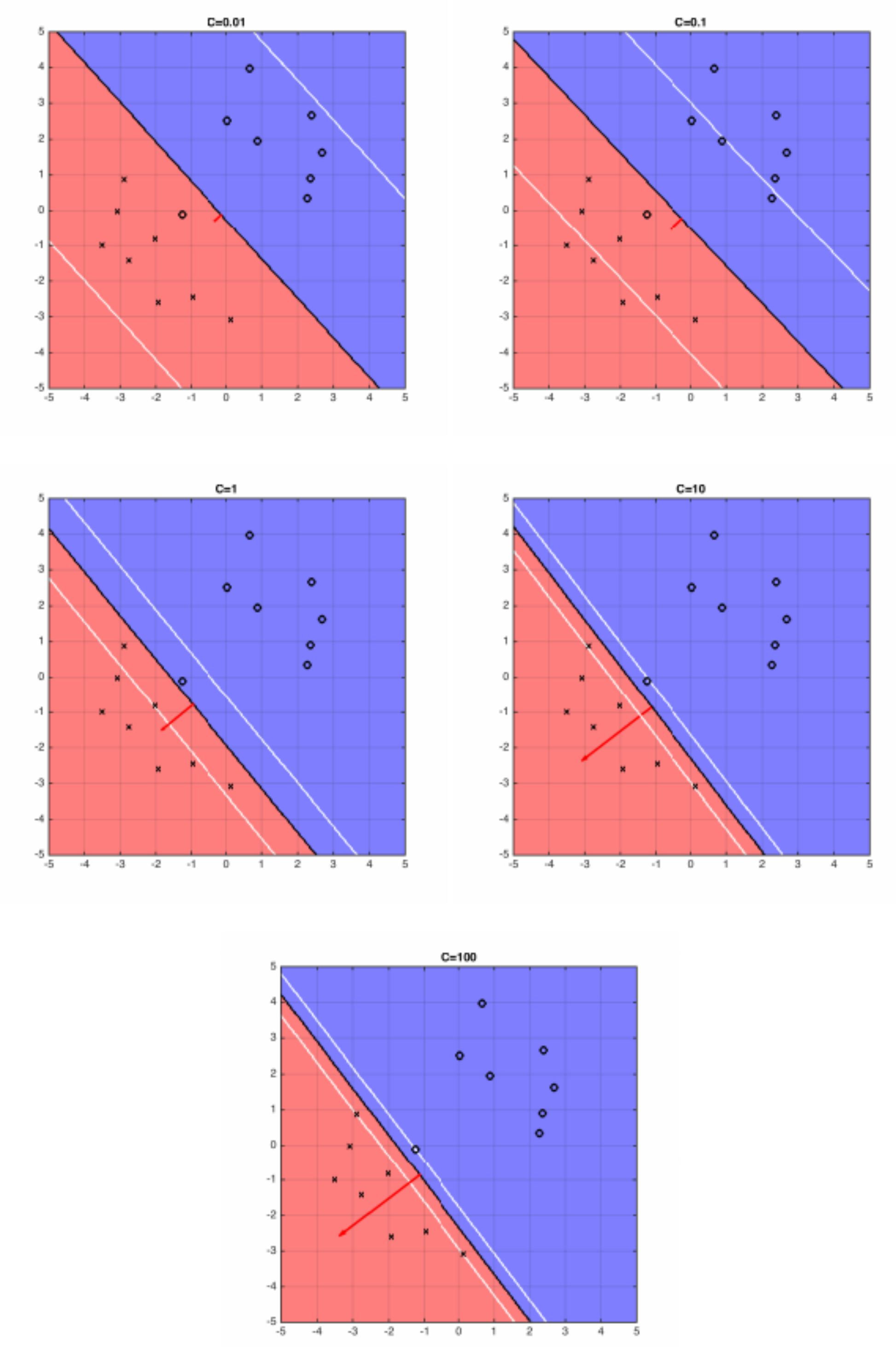


Figure 2: The five plots above show different boundary of hyperplane and the optimal hyperplane separating example data, when C=0.01, 0.1, 1, 10, 100.