

१०५

(1)

(2)

$$= \underset{y}{\arg\max} (\ln (P(x|y) \cdot P(y)))$$

$$\ln(f(x|y)) = -\frac{1}{2} (x - \mu_y)^T \Sigma^{-1} (x - \mu_y) \quad \text{if } x' \in \mathcal{X}$$

$$= -\frac{1}{2} \mu y^T \Sigma^{-1} \mu y + \frac{1}{2} \mu y^T \Sigma^{-1} x \\ + \frac{1}{2} x^T \Sigma^{-1} \mu y - \frac{1}{2} x^T \Sigma^{-1} x$$

$$hUN'O \Sigma_1^{-}$$

④

$$\Rightarrow h_d(x) = \underset{y}{\operatorname{argmax}} (\ln(p(x|y)) + \ln(p(y)))$$

$$= \underset{y}{\operatorname{argmax}} \left(x^t \Sigma^{-1} \mu y - \frac{1}{2} \mu y^t \Sigma \mu y + \ln(p(y)) \right)$$

$$= \underset{y \in \{-1, 1\}}{\operatorname{argmax}} \delta y(x)$$

✓

$$\hat{P}_k^{\text{obs}} = \frac{|\{i : y_i = k\}|}{n} \quad ; \quad (\hat{P}_k = P(Y=k))_{Y \in \mathcal{Y} = \mathcal{L}}$$

$$\hat{\mu}_k^{\text{obs}} = \frac{1}{|\{i : y_i = k\}|} \sum_{i : y_i = k} x_i \quad (k = \pm 1)$$

$$\hat{\Sigma}^{\text{obs}} = \frac{1}{n} \sum_{k=\pm 1} \sum_{i=1}^n (x_i - \hat{\mu}_k) (x_i - \hat{\mu}_k)^T$$

(4) הטבות האפשריות הן להגדיר שמ"ם הוא ספם
אדם הוא לא. או להגדיר שהוא מ"ם ואם אדם
הוא כן ספם
ושגורה מסודר, זו שגורה אדם שנגזר שמ"ם
הוא (אם למרות שהוא) ספם.

positive label \Rightarrow spam
negative label \Rightarrow not spam

3

$$\Rightarrow \operatorname{argmin}_{(w, b)} J(w, b)$$

(5)

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \leq \begin{pmatrix} u_{1 \times 1} & \dots & 1 \\ \vdots & & \vdots \\ u_{m \times m} & & 1 \end{pmatrix} \begin{pmatrix} w \\ b \end{pmatrix} \quad \text{if } w' \text{ is } 1$$

$$\Rightarrow \operatorname{argmin}_{(w, b)} \frac{1}{2} (w, b) (2I) \begin{pmatrix} w \\ b \end{pmatrix} + \sigma^+ \begin{pmatrix} w \\ b \end{pmatrix}$$

$$\begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix} \leq \begin{pmatrix} u_{1 \times 1} & \dots & 1 \\ \vdots & & \vdots \\ u_{m \times m} & & 1 \end{pmatrix} \begin{pmatrix} w \\ b \end{pmatrix} \quad \text{if } w' \text{ is } -1$$

$$\Rightarrow A = \begin{bmatrix} u_{1 \times 1} & \dots & 1 \\ \vdots & & \vdots \\ u_{m \times m} & & 1 \end{bmatrix}$$

$$d = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$$

$$a = 0$$

$$Q = 2I$$

$$\frac{1}{n} \sum_{i=1}^n \epsilon_i$$

(6)

$$\epsilon_i \geq 0 \quad \text{ssi} \quad y_i \langle w, x_i \rangle \geq 1 - \epsilon_i \quad i \in \{1, \dots, n\}$$

$$= \frac{1}{n} \sum_{i=1}^n \epsilon_i$$

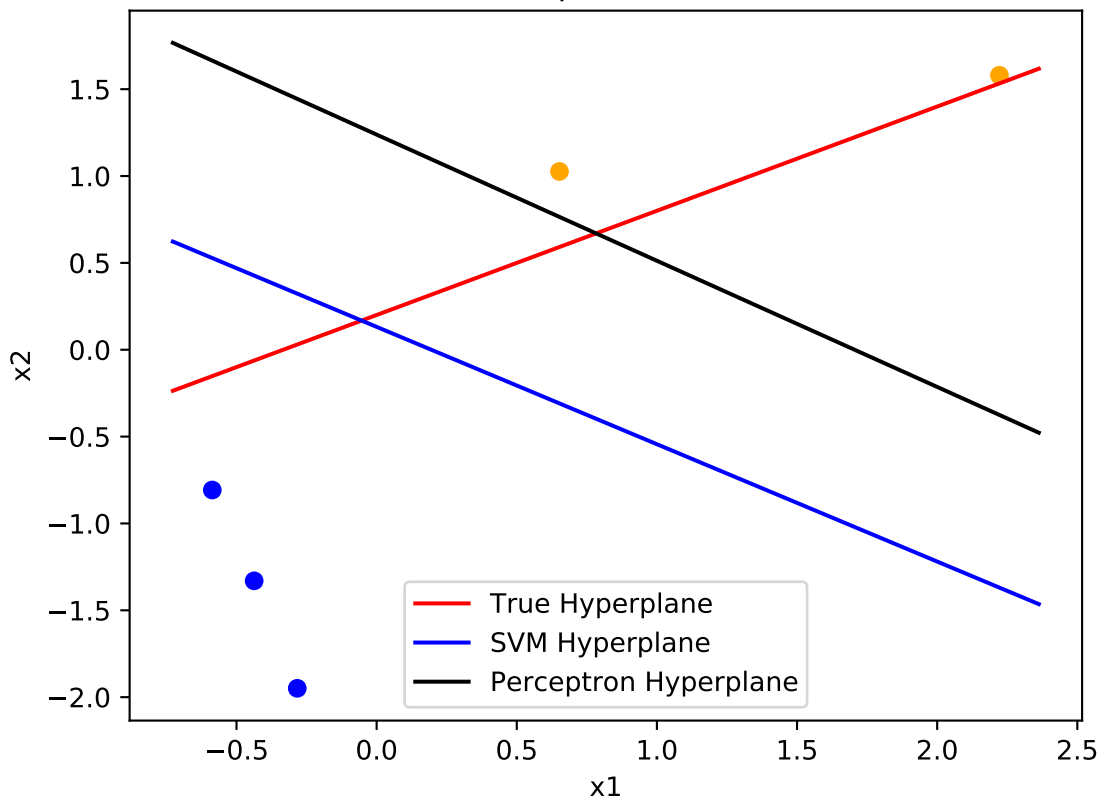
$$\epsilon_i \geq 0 \quad \text{ssi} \quad \epsilon_i \geq 1 - y_i \langle w, x_i \rangle \quad i \in \{1, \dots, n\}$$

$$= \frac{1}{n} \sum_{i=1}^n \max \{0, 1 - y_i \langle w, x_i \rangle\}$$

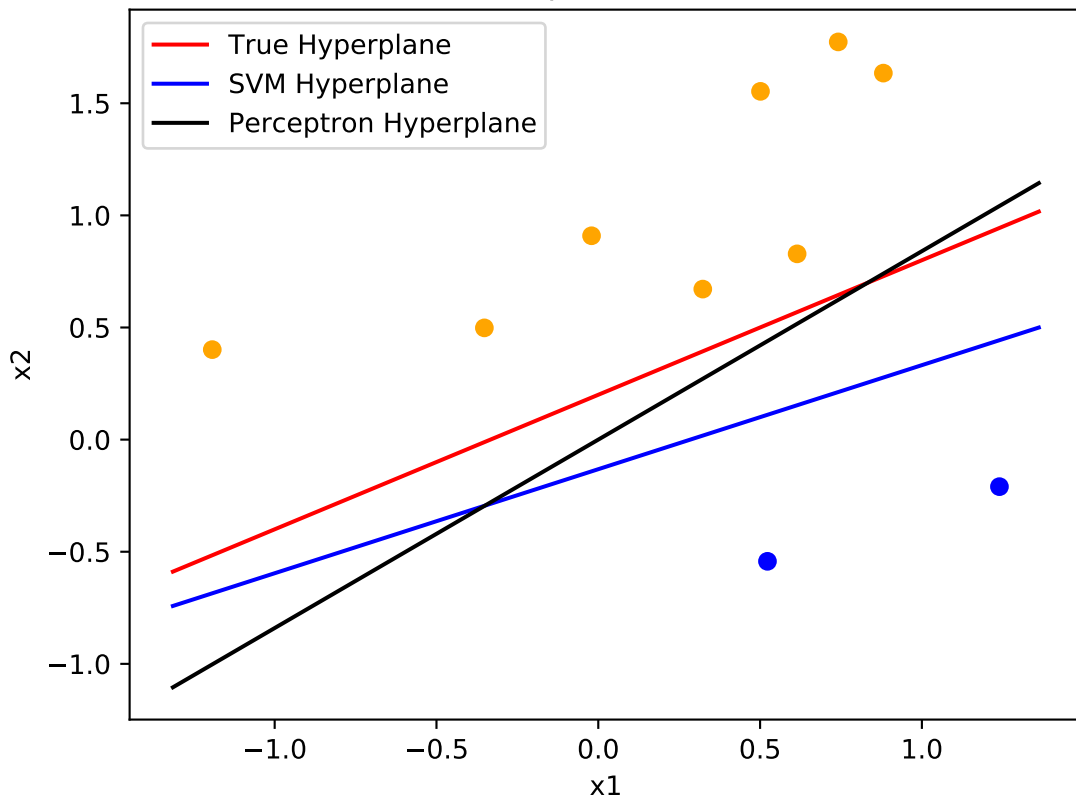
$$\text{on déf } l_{\text{hinge}} = \frac{1}{n} \sum_{i=1}^n l_{\text{hinge}}(y_i \langle w, x_i \rangle)$$

$$\Rightarrow \arg \min \frac{\lambda}{2} \|w\|^2 + \frac{1}{n} \sum_{i=1}^n l_{\text{hinge}}(y_i \langle w, x_i \rangle)$$

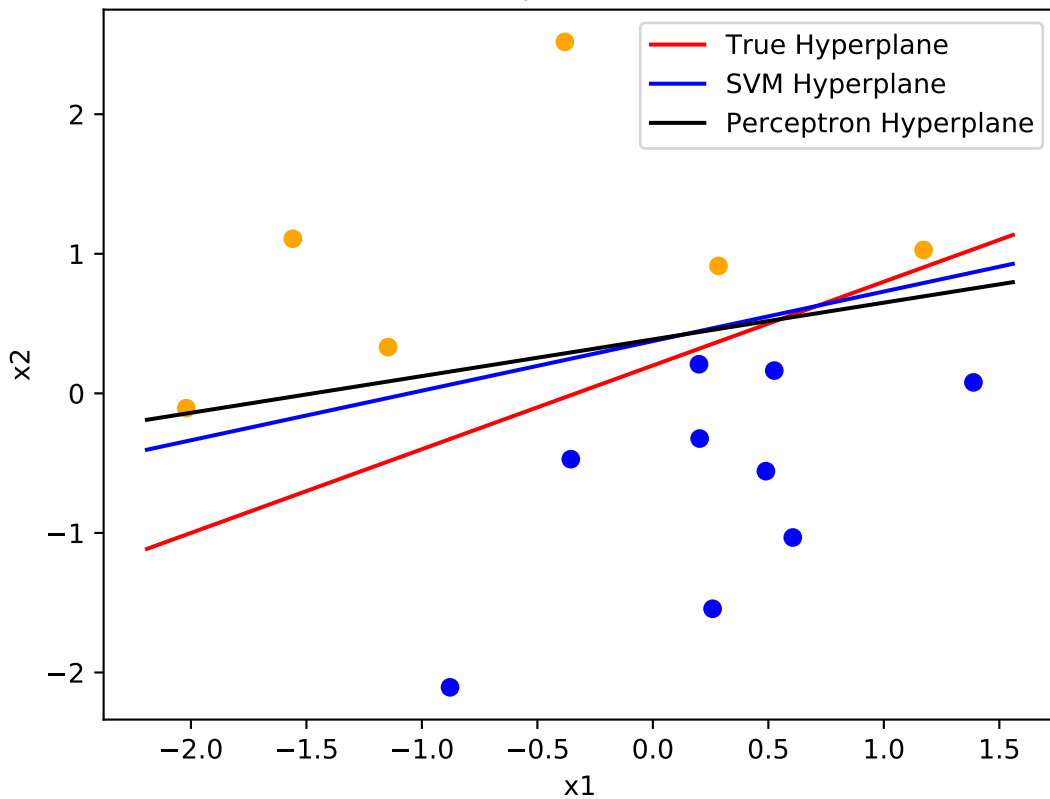
Q9 - Comparaison $m = 5$



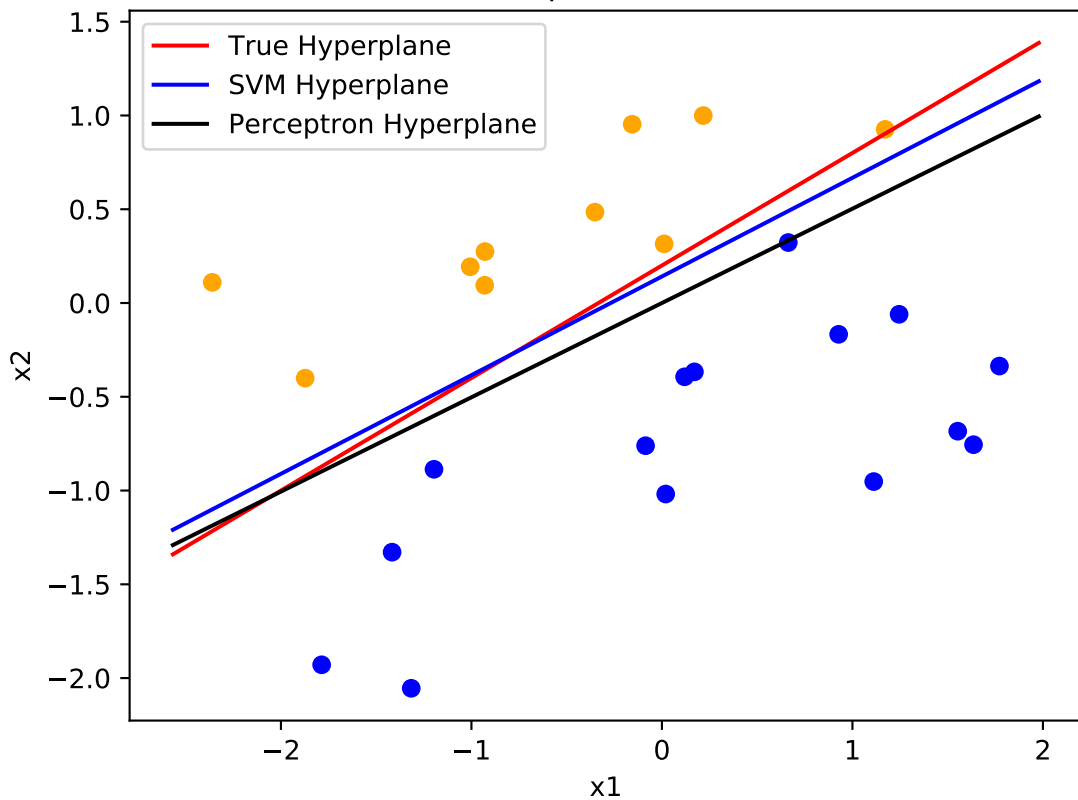
Q9 - Comparaison $m = 10$



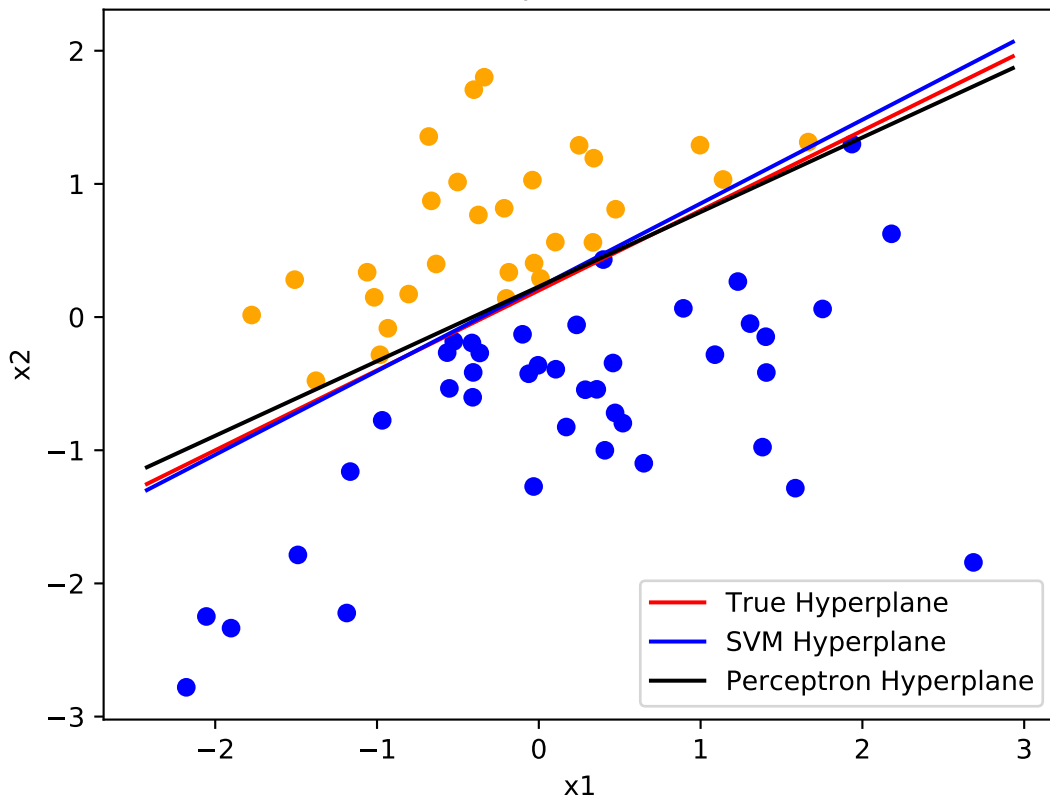
Q9 - Comparaison $m = 15$



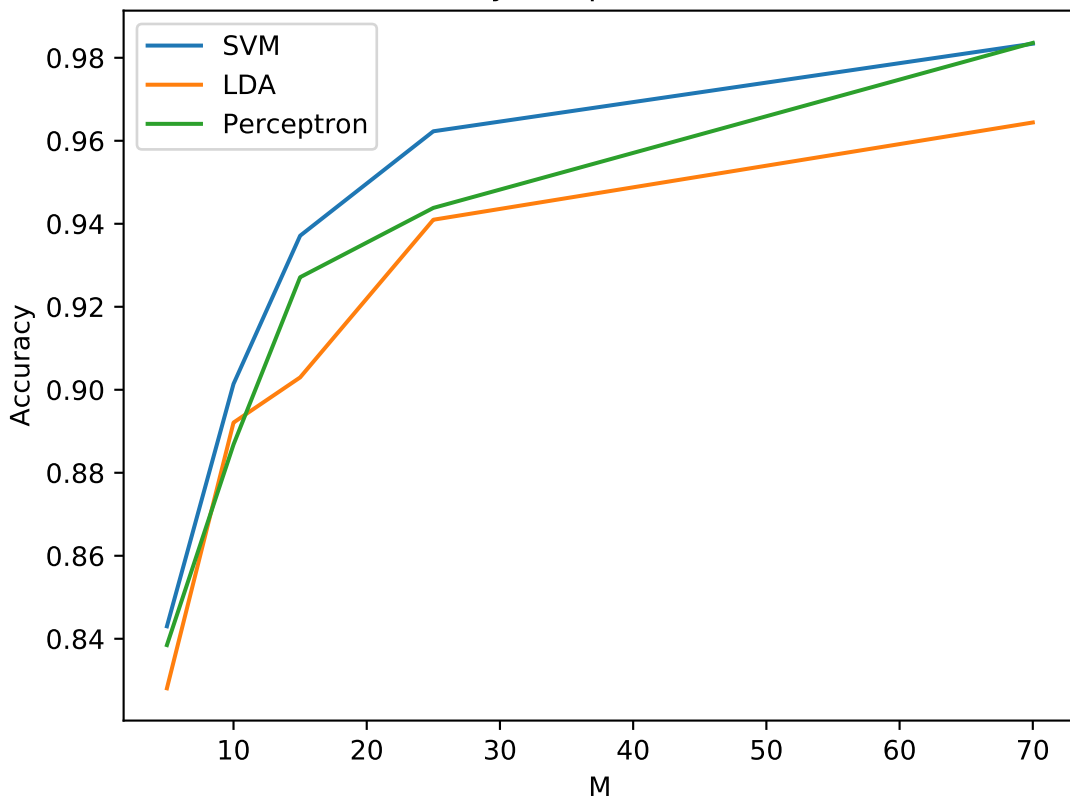
Q9 - Comparaison $m = 25$



Q9 - Comparaison $m = 70$



Q10 - Accuracy comparison of different m



(11) איך החרגת האם מה SVM classifier עשה יותר טוב.

ה-LDA classifier היה עובד יותר טוב עם יותר דאטא (יותר טוב (maximizes margin) SVM עשה יותר טוב.

