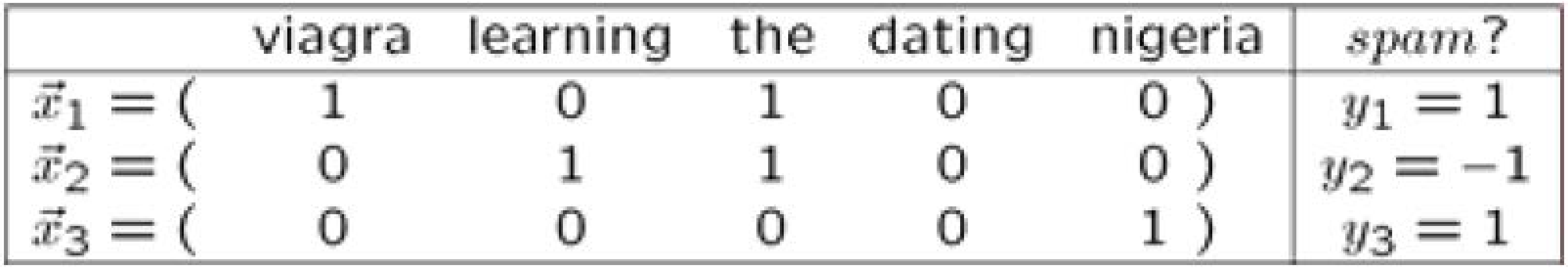
# Large Scale Machine Learning: Support Vector Machines

Mining of Massive Datasets
Leskovec, Rajaraman, and Ullman
Stanford University

# Application: Spam Filtering

- **Example: Spam filtering**



| | viagra | learning | the | dating | nigeria | *spam?* |
|---|---|---|---|---|---|---|
| $\vec{x}_1 = ($ | 1 | 0 | 1 | 0 | 0 ) | $y_1 = 1$ |
| $\vec{x}_2 = ($ | 0 | 1 | 1 | 0 | 0 ) | $y_2 = -1$ |
| $\vec{x}_3 = ($ | 0 | 0 | 0 | 0 | 1 ) | $y_3 = 1$ |

- **Instance space x ∈ X** (|**X**|= **n** data points)
  - **Binary or real-valued feature vector *x* of word occurrences**
  - ***d*** features (words + other things, **d**~100,000)
- **Class y ∈ Y**
  - ***y***: Spam (+1), Ham (-1)

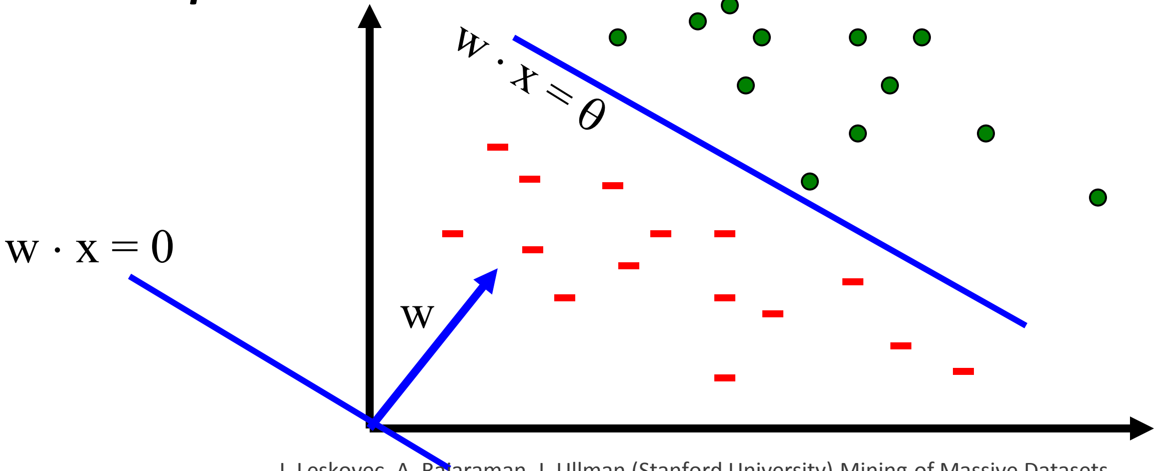
# Linear models for classification

- **Binary classification:**

$$f(\mathbf{x}) = \begin{cases} +1 & \text{if } \mathbf{w}^{(1)}\,\mathbf{x}^{(1)} + \mathbf{w}^{(2)}\,\mathbf{x}^{(2)} + \ldots\, \mathbf{w}^{(d)}\,\mathbf{x}^{(d)} \geq \theta \\ -1 & \text{otherwise} \end{cases}$$

Decision boundary is **linear**

- **Input:** Vectors $\boldsymbol{x}_j$ and labels $\boldsymbol{y}_j$
  - Vectors $\boldsymbol{x}_j$ are binary (real) valued
- **Goal:** Find vector $w = (w^{(1)}, w^{(2)}, \ldots, w^{(d)})$
  - Each $\boldsymbol{w}_i$ is a real number



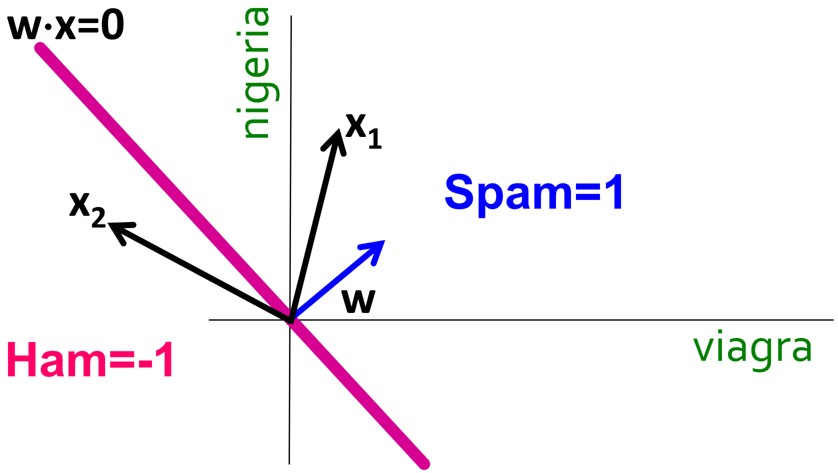
**Note:**

$$\mathbf{x} \Leftrightarrow \langle \mathbf{x}, 1 \rangle \quad \forall \mathbf{x}$$

$$\mathbf{w} \Leftrightarrow \langle \mathbf{w}, -\theta \rangle$$

# Linear Classifiers

- Each feature has a weight $w^{(i)}$
- **Prediction is based on the weigthed sum:**
  - $f(x) = \Sigma_i \, w^{(i)} \, x^{(i)} = w \cdot x$
- If the ***f(x)*** is:
  - **Positive:** Predict **+1**
  - **Negative:** Predict **-1**


w·x=0
nigeria
x₁
Spam=1
x₂
w
Ham=-1
viagra