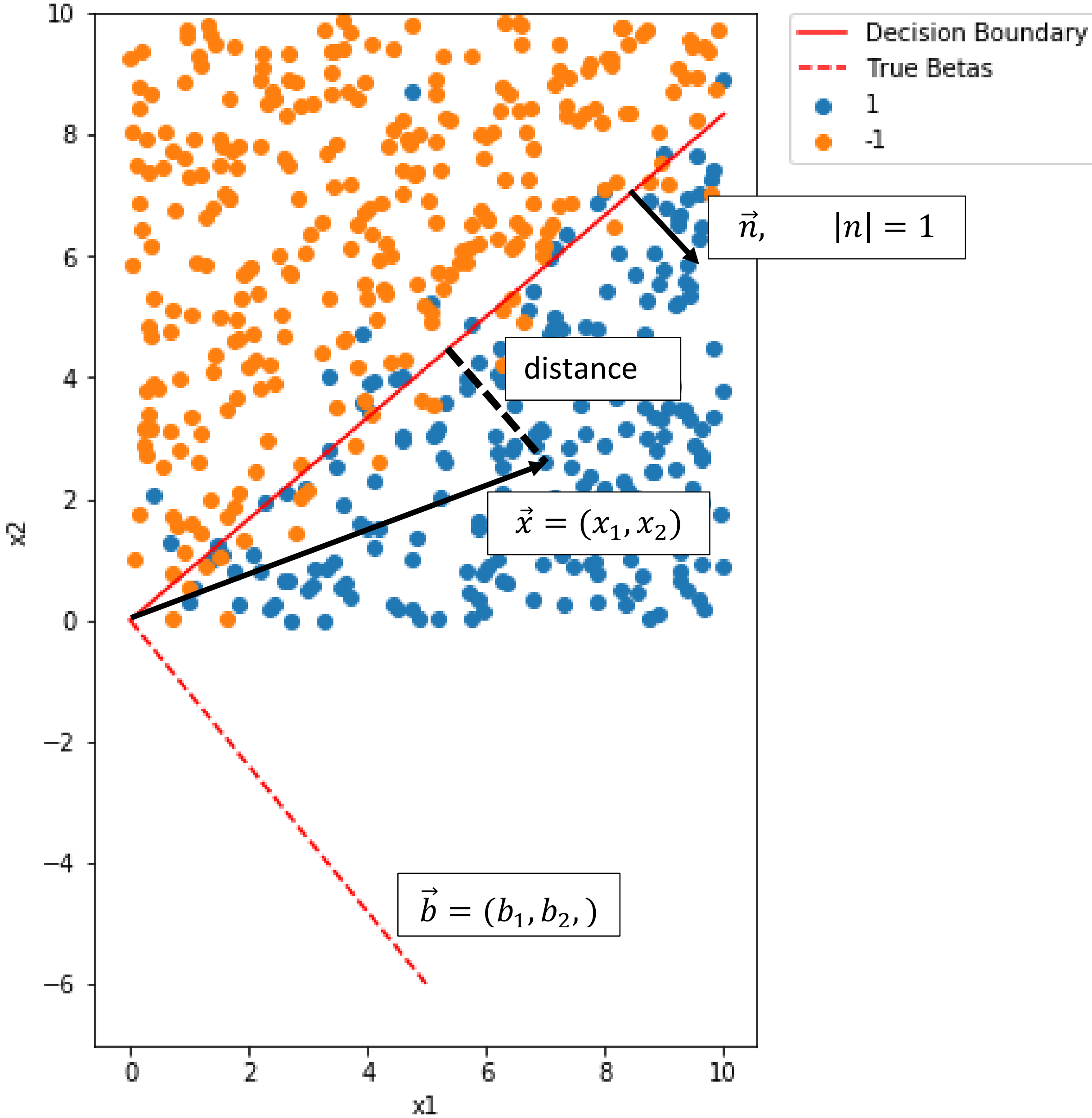

Probability of sample to be 1 depends on distance d:

$$P(y = 1|x, b) = f(d)$$

$$d = \vec{x}\,\vec{n}, \quad \vec{n}\ is\ normal\ of\ Decision\ Boundary$$

Assume:

$$f(d) = \frac{1}{1 + e^{-\alpha d}}, \quad \alpha \; is \; any \; constant$$

$$= \frac{1}{1 + e^{-\alpha \vec{x}\vec{n}}} = \frac{1}{1 + e^{-\vec{b}\vec{x}}}, \quad where \; \vec{b} = \alpha\vec{n}$$

So:

$$P(y = 1|x, b) = \frac{1}{1 + e^{-\vec{b}\vec{x}}}$$

$$P(y = -1|x, b) = 1 - P(y = 1 \,\middle|\, x, b) = \frac{1}{1 + e^{\vec{b}\vec{x}}}$$

Combining two equations above:

$$P(y|x, b) = \frac{1}{1 + e^{-y\vec{b}\vec{x}}}$$

Probability that we predict y_t for t-th sample (**i.e. make true prediction**) is:

$$P(y_t|x_t, b) = \frac{1}{1 + e^{-y_t\vec{b}\vec{x}_t}}$$

Probability that we make true predictions for all samples is:

$$P(Y|X, b) = \prod_t \frac{1}{1 + e^{-y_t\vec{b}\vec{x}_t}}$$

We want this value to be as big as possible, so we should find $\vec{b}^*$ which maximize $P(Y|X, b)$:

$$\vec{b}^* = \max_{\vec{b}} \prod_t \frac{1}{1 + e^{-y_t\vec{b}\vec{x}_t}} =$$

$$= \max_{\vec{b}} \sum_t \log \frac{1}{1 + e^{-y_t\vec{b}\vec{x}_t}} = \min_{\vec{b}} \sum_t \underbrace{\log\left(1 + e^{-y_t\vec{b}\vec{x}_t}\right)}_{l_t(\vec{b})}$$

To resolve this problem, we use minimization method "Stochastic Gradient Descend":

1 step:

Take $\vec{b} = (0, 0)$

2 step:

Find the gradient of log-loss for first (next) sample:

$$\frac{d}{d\vec{b}} l_t(\vec{b}) = -y_t \vec{x}_t \frac{e^{-y_t \vec{b} \vec{x}_t}}{1 + e^{-y_t \vec{b} \vec{x}_t}}$$

3 step:

Update $\vec{b}$ to opposite direction of gradient:

$$\vec{b} = \vec{b} + \eta y_t \vec{x}_t \frac{e^{-y_t \vec{b} \vec{x}_t}}{1 + e^{-y_t \vec{b} \vec{x}_t}}$$

Go to step 2 and repeat for all date