	Supervised Learning	*Unsupervised Learning*
Discrete	classification or categorization Discrete Output $y \in \{1, \ldots, K\}$	clustering
Continuous	regression Continuous Output $y \in R$	dimensionality reduction

For data X and class Y

GENERATIVE

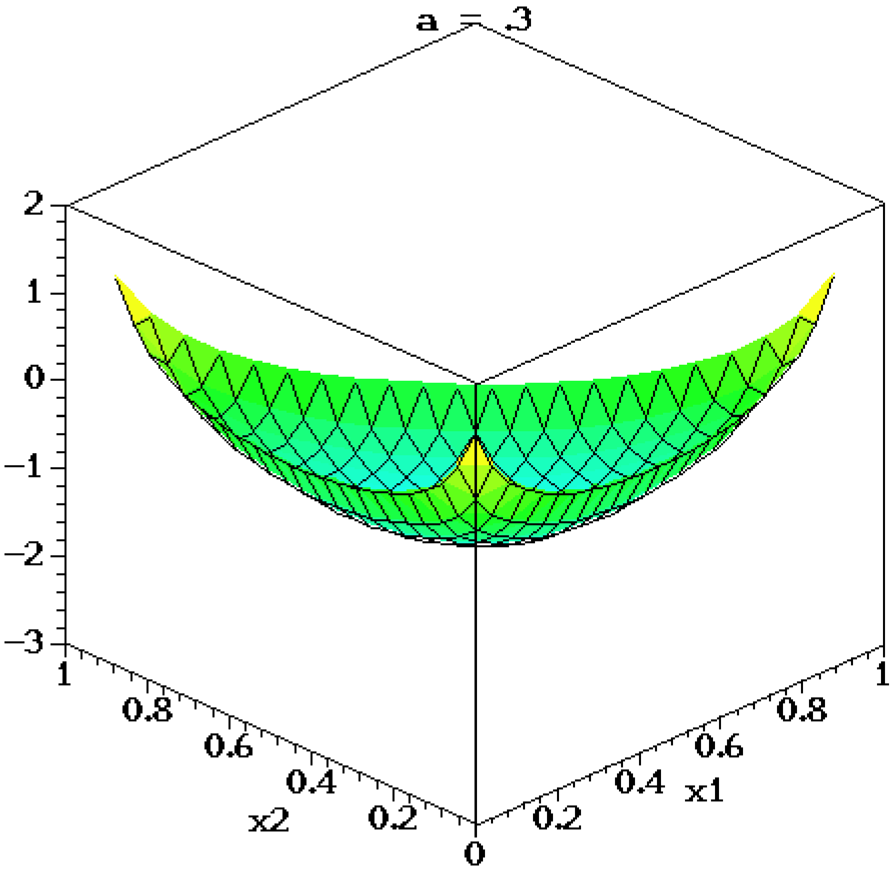
$$P(Y,X) = P(X|Y)\,P(Y)$$

$$P(Y|X) = \frac{P(Y,X)}{P(X)}$$

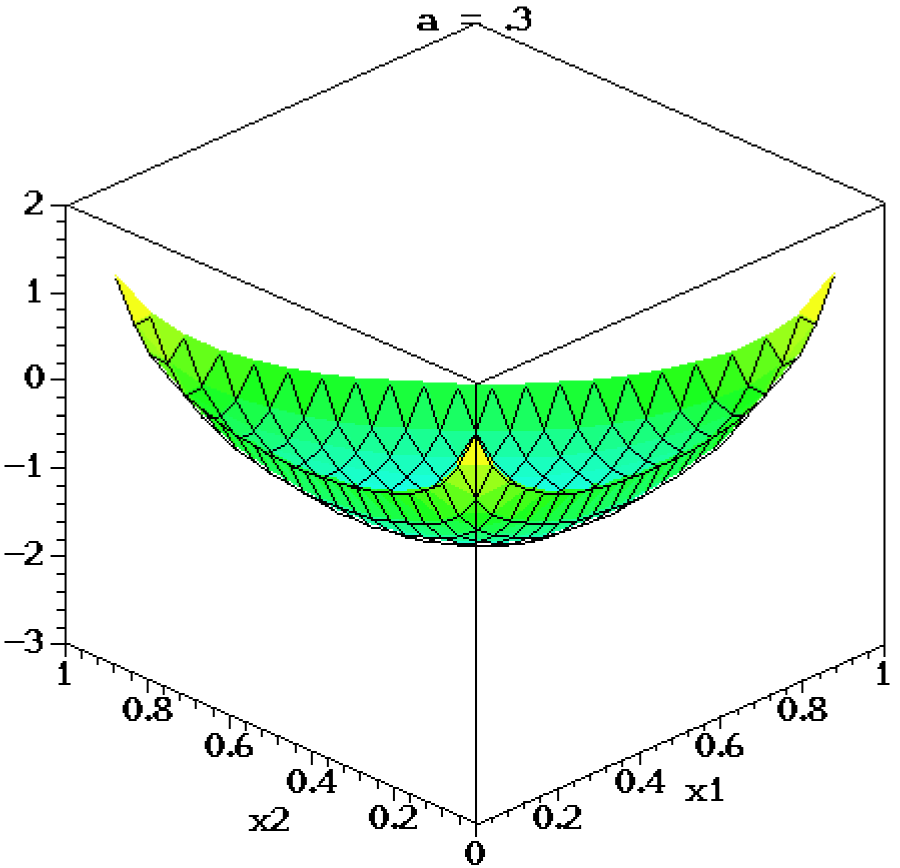
DISCRIMINATIVE

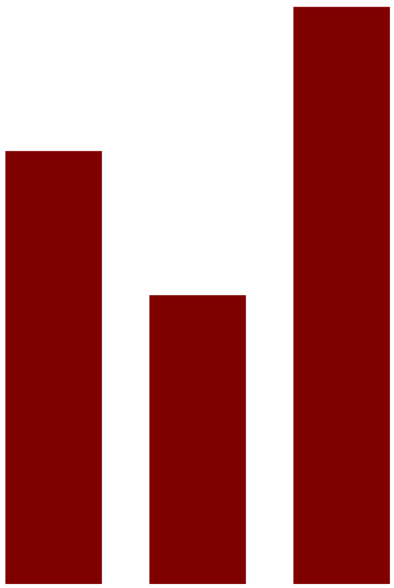
$$P(Y|X)$$

DIRICHLET



DIRICHLET MULTINOMIAL

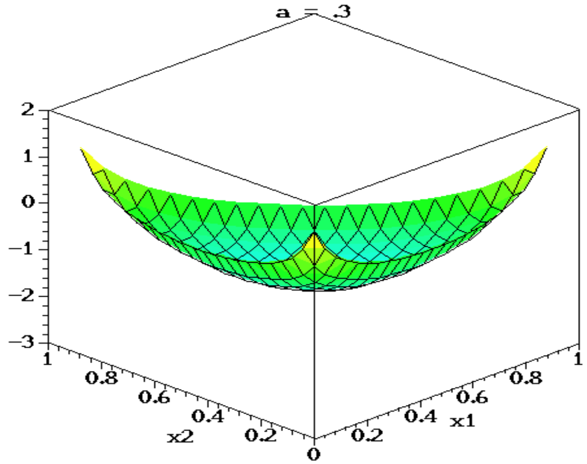


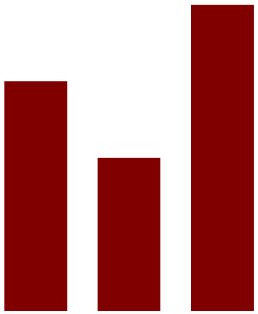


DIRICHLET

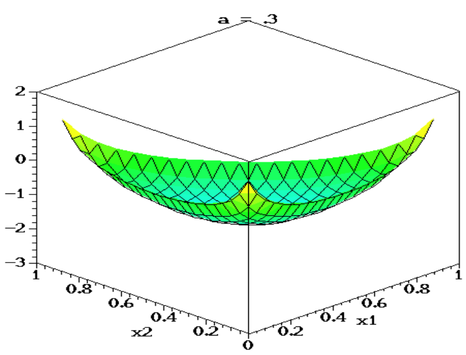
MULTINOMIAL

PRIOR

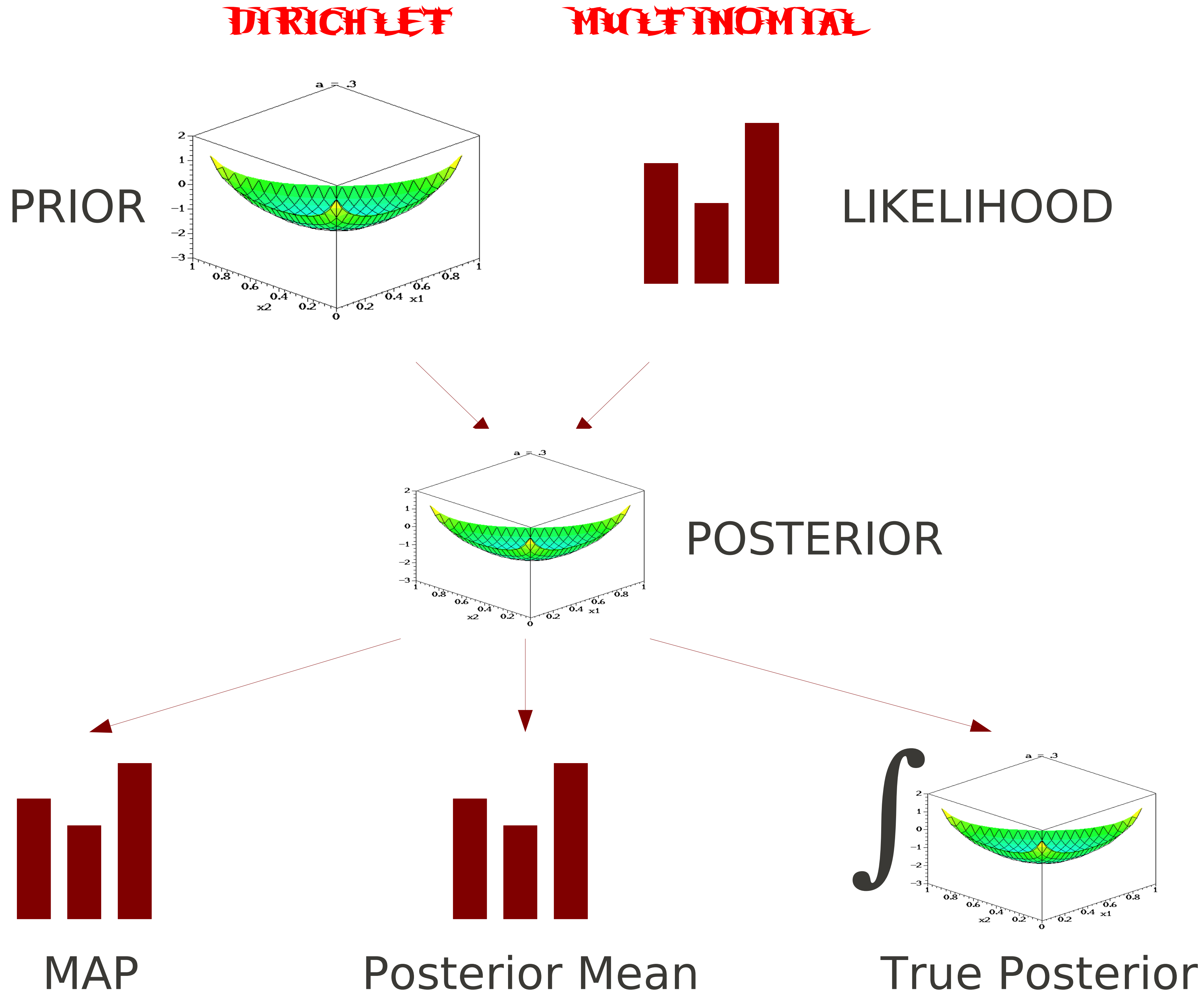




LIKELIHOOD



POSTERIOR


DIRICHLET
MULTINOMIAL
PRIOR
a = .3
x1
x2
LIKELIHOOD
POSTERIOR
MAP
Posterior Mean
True Posterior

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

$$Dir(\alpha) = C\prod_{i=1}^{K}\theta_i^{\alpha_i - 1}$$

$$Multi(x|\theta) = C\prod_{i=1}^{K}\theta_i^{x_i}$$

- In Bayesian approach we minimize *posterior expected loss*

$$\rho(\mathbf{a}|\mathbf{x},\pi) := \mathbb{E}_{p(\boldsymbol{\theta}|\mathbf{x},\pi)}\left[L(\boldsymbol{\theta},\mathbf{a})\right] = \int_{\Theta} L(\boldsymbol{\theta},\mathbf{a})p(\boldsymbol{\theta}|\mathbf{x},\pi)d\theta$$

- MAP estimate minimizes 0-1 loss:

$$L(\theta, a) = \mathbb{I}(\theta \neq a) = \begin{cases} 0 & \text{if } a = \theta \\ 1 & \text{if } a \neq \theta \end{cases}$$

- Plugging into $\rho(a|\mathbf{x})$ yields

$$\rho(a|\mathbf{x}) = p(a \neq y|\mathbf{x}) = 1 - p(a{=}y|x)$$

$$y^*(\mathbf{x}) = \arg\max_{y \in \mathcal{Y}} p(y|\mathbf{x}), \qquad (MAP)$$

- Posterior mean minimizes L2 loss,

$$L(\theta, a) = (\theta - a)^2$$

- Plug into expected loss,

$$\rho(a|\mathbf{x}) \quad = \quad \mathbb{E}\left[(\theta - a)^2|\mathbf{x}\right] = \mathbb{E}\left[\theta^2|\mathbf{x}\right] - 2a\mathbb{E}\left[\theta|\mathbf{x}\right] + a^2$$

$$\frac{\partial}{\partial a}\rho(a|\mathbf{x}) = -2\mathbb{E}\left[\theta|\mathbf{x}\right] + 2a = 0$$

$$\Rightarrow \; a = \mathbb{E}\left[\theta|\mathbf{x}\right] = \int \theta p(\theta|\mathbf{x}) d\theta$$

- Posterior median minimizes L1 loss

- Consider *loss matrix*

	$\hat{y}=1$	$\hat{y}=0$
$y=1$	0	L_{FN}
$y=0$	L_{FP}	0

, L_{FN} – *False Negative*
, L_{FP} – *False Positive*

- Posterior expected loss is

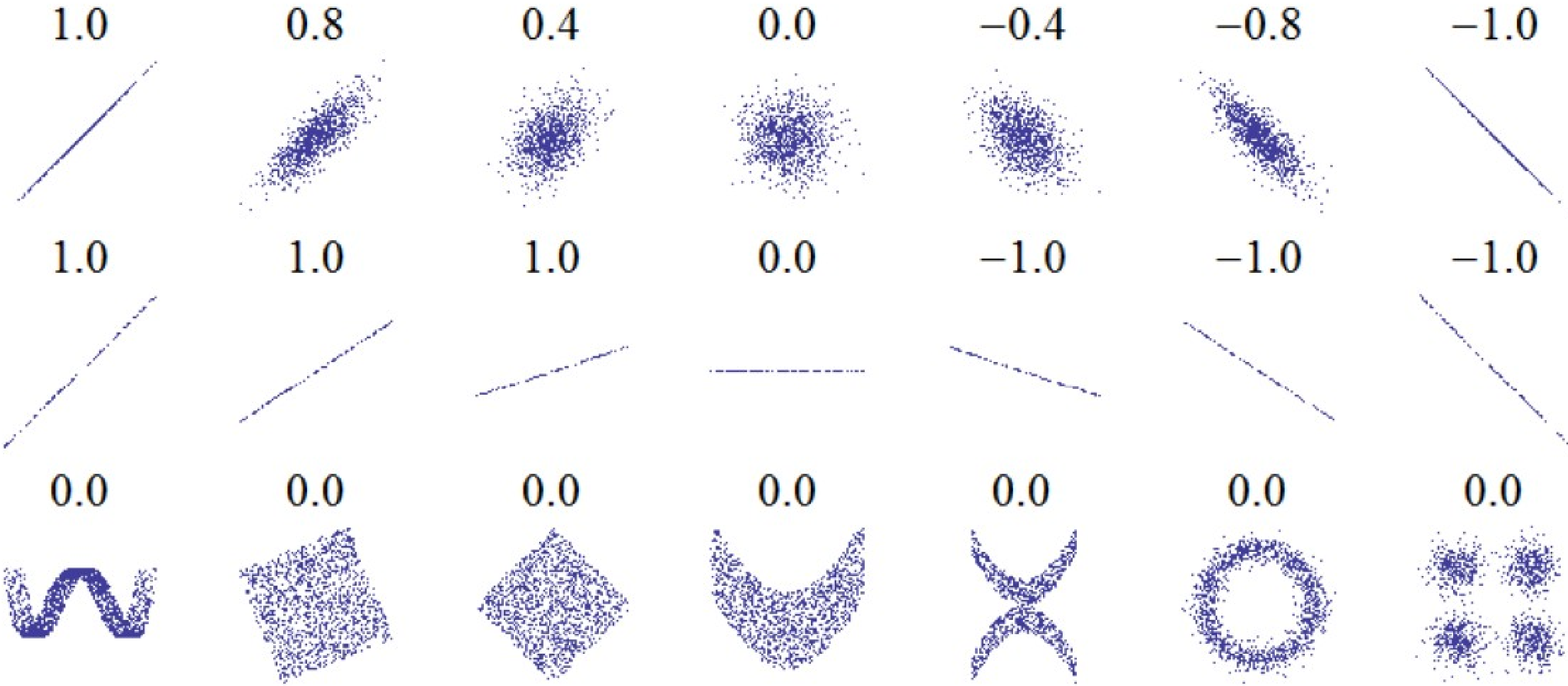
$$\rho(\hat{y}=0|\mathbf{x}) = L_{FN} \times p(y=1|\mathbf{x})$$
$$\rho(\hat{y}=1|\mathbf{x}) = L_{FP} \times p(y=0|\mathbf{x})$$

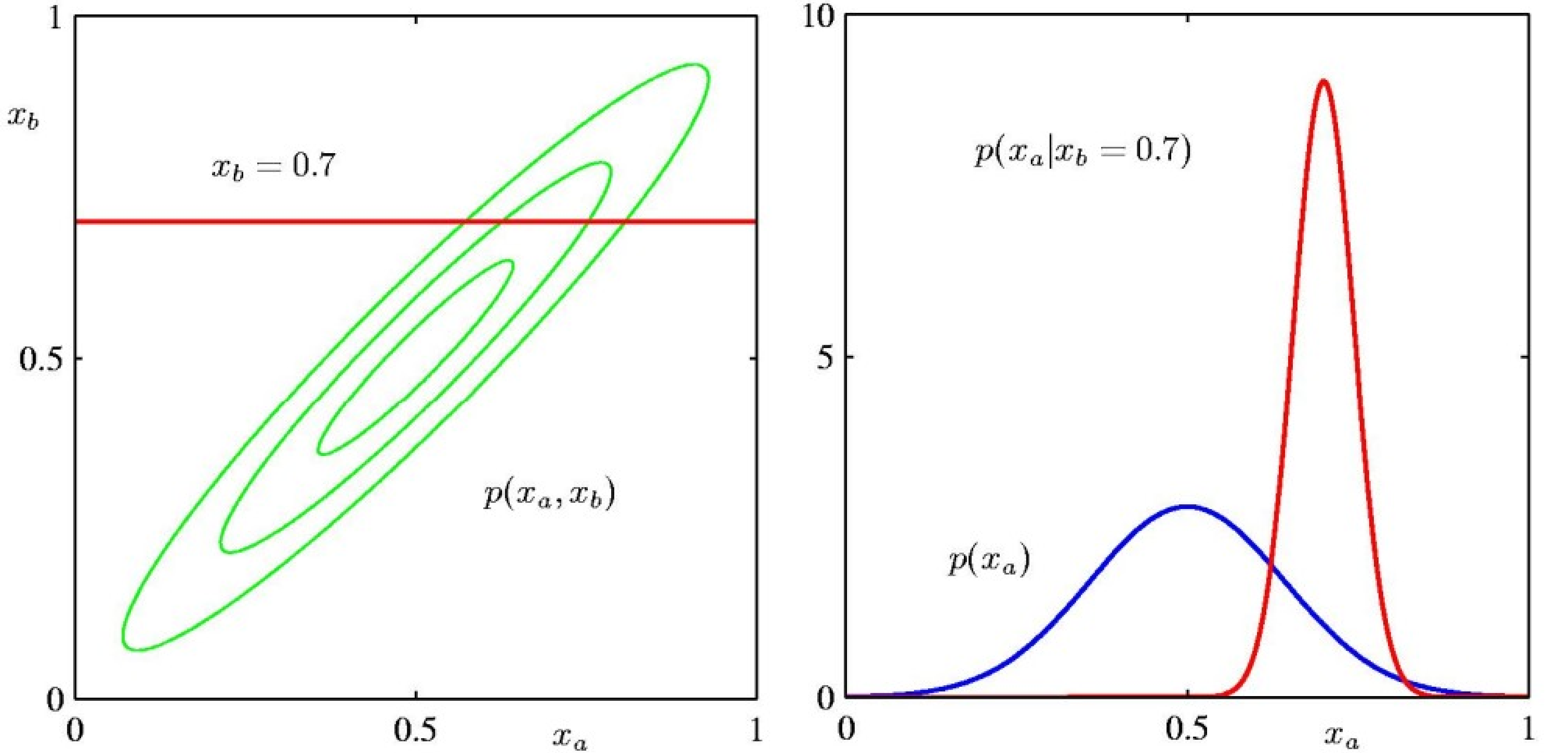
- So we pick class $\hat{y}=1$ iff

$$\rho(\hat{y}=0|\mathbf{x}) > \rho(\hat{y}=1|\mathbf{x})$$
$$\frac{p(y=1|\mathbf{x})}{p(y=0|\mathbf{x})} > \frac{L_{FP}}{L_{FN}} = \eta$$

Can trace out ROC curve
By varying η

$$\operatorname{cov}[X,Y] \triangleq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY]-\mathbb{E}[X]\,\mathbb{E}[Y]$$





1
x_b
$x_b = 0.7$
0.5
$p(x_a, x_b)$
0
0
0.5
x_a
1
10
$p(x_a|x_b = 0.7)$
5
$p(x_a)$
0
0
0.5
x_a
1