EXPONENTIAL FAMILIES & GLMs

$y|x;\theta \sim N(\boxed{\theta^T x}, \sigma^2)$ Regression $y \in \mathbb{R}$

$y|x;\theta \sim \mathrm{Bern}\left(\frac{1}{1+e^{\boxed{-\theta^T x}}}\right)$ Classification $y \in \{0,1\}$

pdf / pmf

$$p(y;\eta) = b(y)\exp\{\eta^T T(y) - a(\eta)\}$$

Most cases

$b(y)$ – Base Measure

$\boxed{y} = T(y)$ – Sufficient Statistic

$a(\eta)$ – log partition function

$\eta \rightarrow$ natural parameter

$\rightarrow$ Probability distribution

$$p(y;\eta) \propto b(y)e^{\eta^T y}$$

$$p(y;\eta) = \frac{b(y)e^{\eta^T y}}{\int b(y)e^{\eta^T y}\,dy} = \frac{b(y)e^{\eta^T y}}{E[e^{\eta^T y}]}$$

M.G.F.

$$= b(y)\, e^{\eta^T y - \log A(\eta)}$$

$$\boxed{a(\eta) = \log A(\eta)}$$

Bernoulli

$$P(y;\phi) = \phi^{y}(1-\phi)^{1-y}$$

$$= \exp\left[\log\left(\phi^{(y)}(1-\phi)^{(1-y)}\right)\right]$$

$$= \exp\left[y\log\phi + (1-y)\log(1-\phi)\right]$$

$$= \exp\left[y\log\phi - y\log(1-\phi) + \log(1-\phi)\right]$$

$$= \exp\left[y\log\left(\frac{\phi}{1-\phi}\right) + \log(1-\phi)\right]$$

$$P(y;\eta) = b(y)\exp\{\eta^T y - a(\eta)\}$$

$$\boxed{\eta = \log\left(\frac{\phi}{1-\phi}\right)} \qquad \boxed{T(y) = y}$$

$$\phi = \frac{1}{1+e^{-\eta}} \rightarrow \text{Logistic function}$$

$$a(\eta) = -\log(1-\phi)$$
$$= \log(1+e^{\eta})$$

$$b(y) = 1$$

Gaussian

$$p(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2} \right\}$$

Assume $\sigma^2 = 1$

$$p(y;\mu) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}(y-\mu)^2 \right\}$$

$$= \left[\frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} y^2 \right\} \right] \exp\left\{ \mu y - \frac{1}{2}\mu^2 \right\}$$

$$p(y;\eta) = b(y) \exp\left\{ \eta^T T(y) - a(\eta) \right\}$$

$\eta = \mu$ $\quad a(\eta) = +\frac{1}{2}\mu^2$

$T(y) = y$ $\quad = \frac{1}{2}\eta^2$

$$b(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\}$$

↳ Standard Gaussian

$\mu = 0$

$\sigma = 1$

<1> $\log P(y;\eta)$ is concave in η

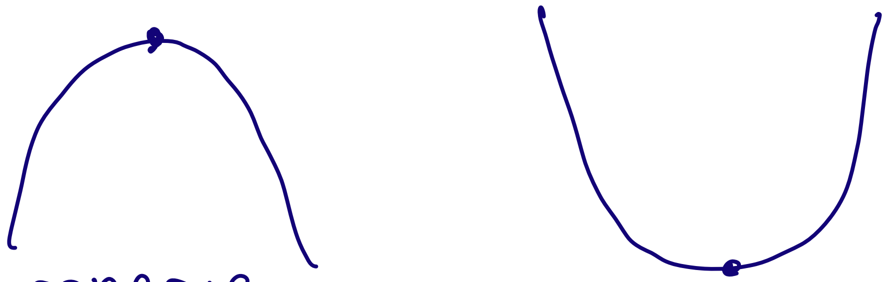
MLE → concave in η

NLL → convex in η.

<2> $$E[y;\eta] = \frac{\partial}{\partial \eta} a(\eta)$$

<3> $$var[y;\eta] = \frac{\partial^2 a(\eta)}{\partial^2 \eta}$$

These are especially nice because when given distribution, to find E, you have to integral but these you have differentiation.



concave

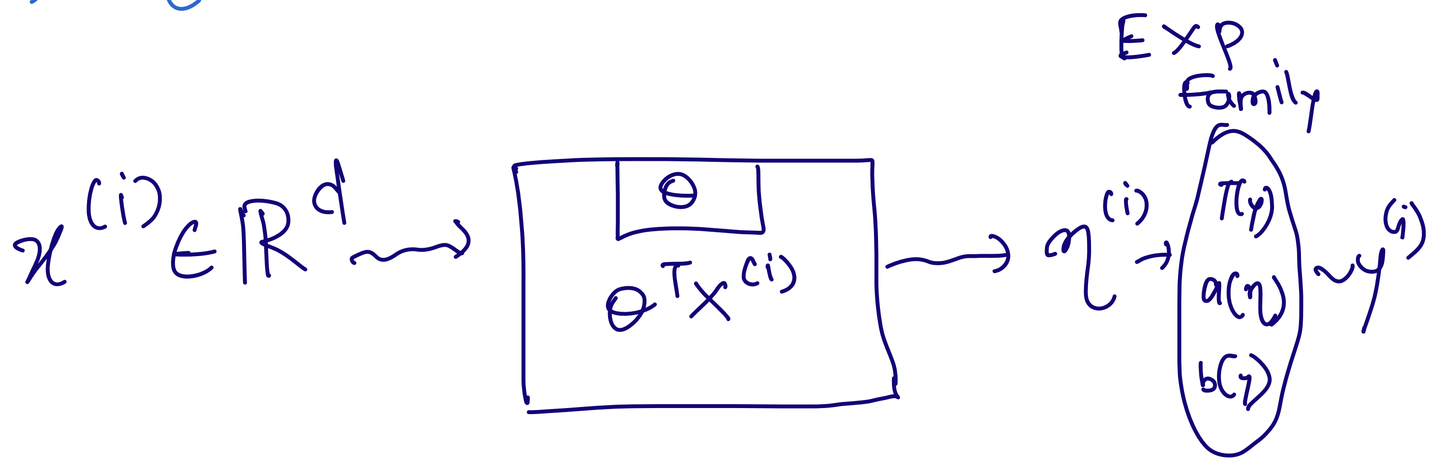
convex

G.L.M.

<1> $y|x;\theta \sim \text{ExpFamily}(\eta)$

<2> $h_\theta(x) = E[y|x;\theta]$

<3> $\eta = \theta^T x$



We first fix a distribution which fixes $T(y), a(\eta), b(y)$ which models our output

$\theta \rightarrow$ global

$x^{(i)} \rightarrow$ local

$\eta = \theta^T x =$ local

LINEAR REGRESSION

Exp Family → Gaussian

$$h_\theta(x) = E(y \mid x; \theta)$$

$$= \mu$$

$$= \eta$$

$$= \theta^T x$$

LOGISTIC

Exp family is Bernoulli

$$h_\theta(x) = E(y|x;\theta)$$

$$= \phi$$

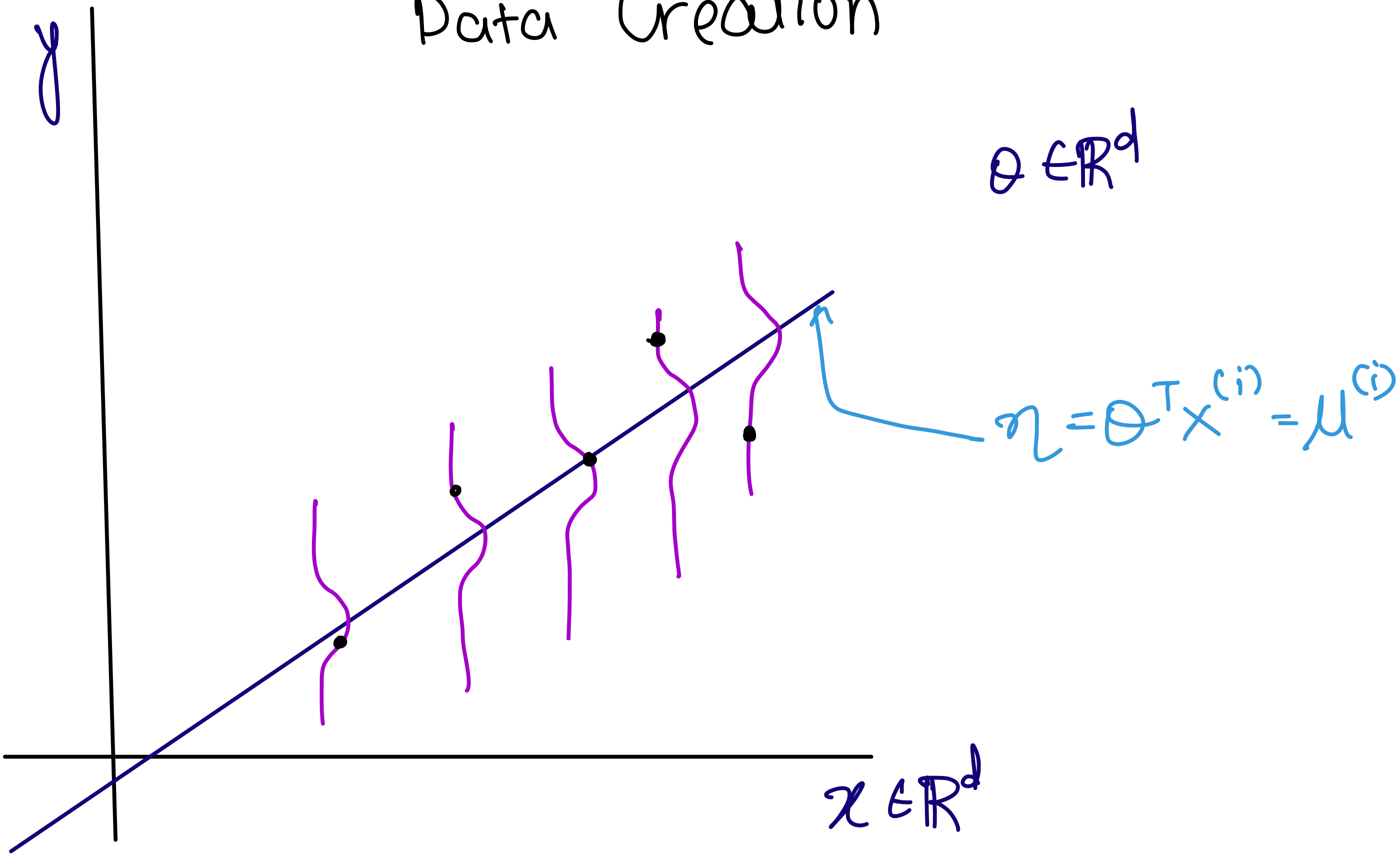
$$= \frac{1}{1+e^{-\eta}}$$

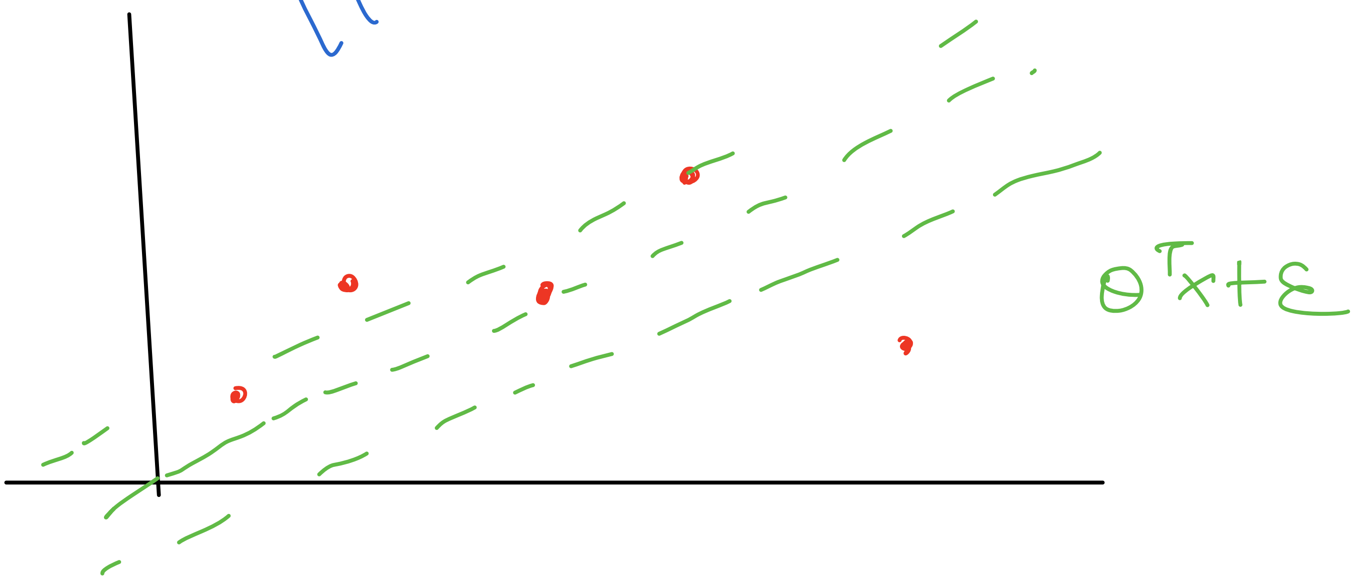
$$= \frac{1}{1+e^{-\theta^T x}}$$

$$= g(\theta^T x)$$

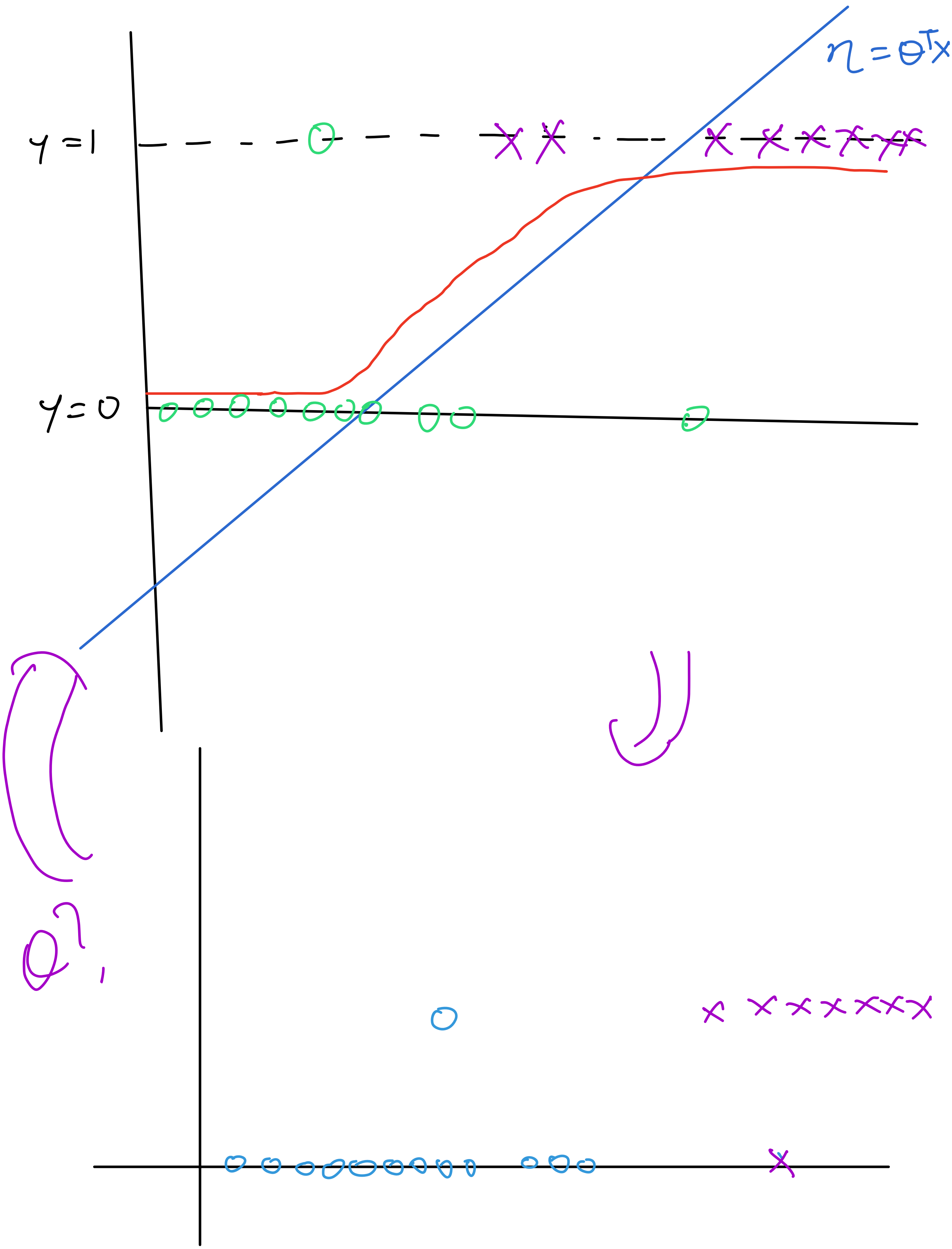
$$g(z) = \frac{1}{1+e^{-z}}$$

Data Creation





We assume first that data generated according to Gaussian. Then when we have data, we actually don't now θ, we assume there is true θ & work to find it $\hat{\theta}$ that is close to true θ.



y=1
y=0
η=θ^T x
θ^T,

MODEL | NATURAL | MEAN

$$\theta \in \mathbb{R}^d,\quad x^{(i)} \in \mathbb{R}^d \xrightarrow{\theta^T x^{(i)}} \eta^{(i)} \xrightarrow{g} \begin{cases} \mu & \text{Gaussian} \\ \phi & \text{Bernoulli} \\ \lambda & \text{poisson} \end{cases}$$

→ Canonical Response function

Gaussian $g(\eta) = \eta$

Bern $\phi = g(\eta) = \dfrac{1}{1+e^{-\eta}}$

$g^{-1} \rightarrow$ Canonical link function

y – response function

Prediction : $h_\theta(x) = E[y \mid x; \theta] = g(\theta^T x)$

G.L.M.

Data Type	Exp Fam Dist^n	Name
$\mathbb{R}$	Gaussian Laplace	Regression
$\{0, 1\}$	Bernoulli	Classification
$\{1 \dots K\}$	Categorical	Multiclass Classification
N_+	Poisson	Count Reg/ Poisson Reg
R_+ (time)	Exponential, Gamma	Survival Analysis

<1> Make choice of dist according to data type

<2> Express in exp. form: - $a(\eta)$, $b(y)$, $T(y)$

$\mu, \phi = g(\eta)$

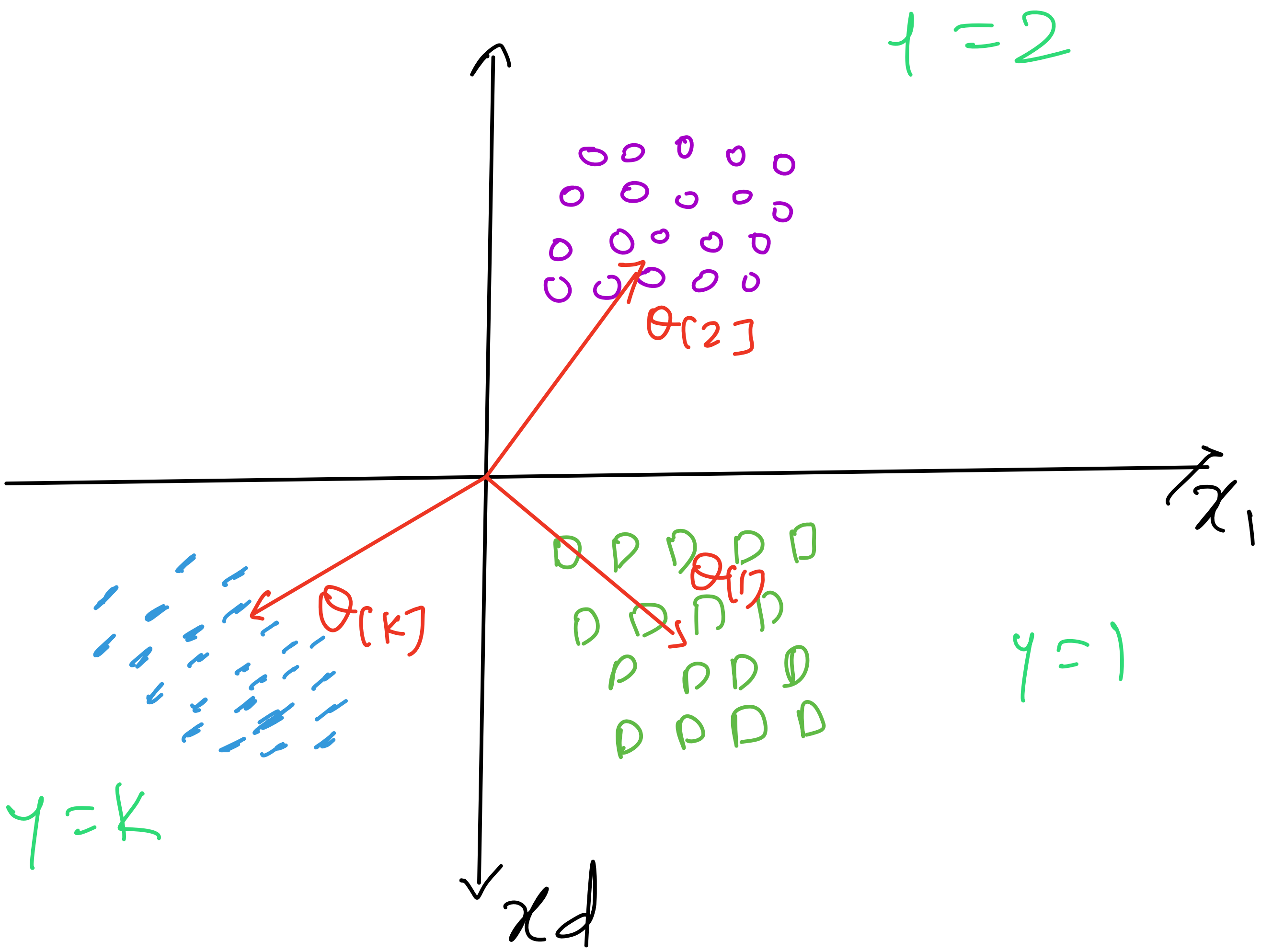
<3> Hypothesis $h_\theta(x) = g(\theta^T X)$

<4> $\theta := \theta + \alpha\left(y^{(i)} - h_\theta(x^{(i)})\right) . x^{(i)}$

<5> Prediction :- $\hat{y} = h_\theta(x^*) = g(\theta^T x^*)$

SOFTMAX REGRESSION

Multiclass Classification



One θ per class

$$\left.\begin{array}{ll} \theta_{[1]}^T x & - \mathbb{R} \\ \theta_{[2]}^T x & - \mathbb{R} \\ \vdots & \\ \theta_{[k]}^T x & - \mathbb{R} \end{array}\right\} \underset{\theta}{\arg\max}$$

$$\theta = \underbrace{\begin{bmatrix} - \theta_{[1]}^T - \\ \vdots \\ - \theta_{[k]}^T - \end{bmatrix}}_{d} \updownarrow k \quad \begin{bmatrix} \\ x \\ \vdots \\ \end{bmatrix} \updownarrow d \quad \begin{array}{c} \theta_{[1]}^T x \\ \theta_{[2]}^T x \\ \vdots \\ \theta_{[k]}^T x \end{array}$$

Then exp $\theta^T x$ & normalise it

$$\frac{\exp(x^T \theta_{[i]})}{\sum_j \exp(x^T \theta_{[j]})}$$

Maximum component of $\vec{x}$

$\rightarrow \quad \vec{x}^T \text{softmax}(\vec{x}) \approx \max(\vec{x})$