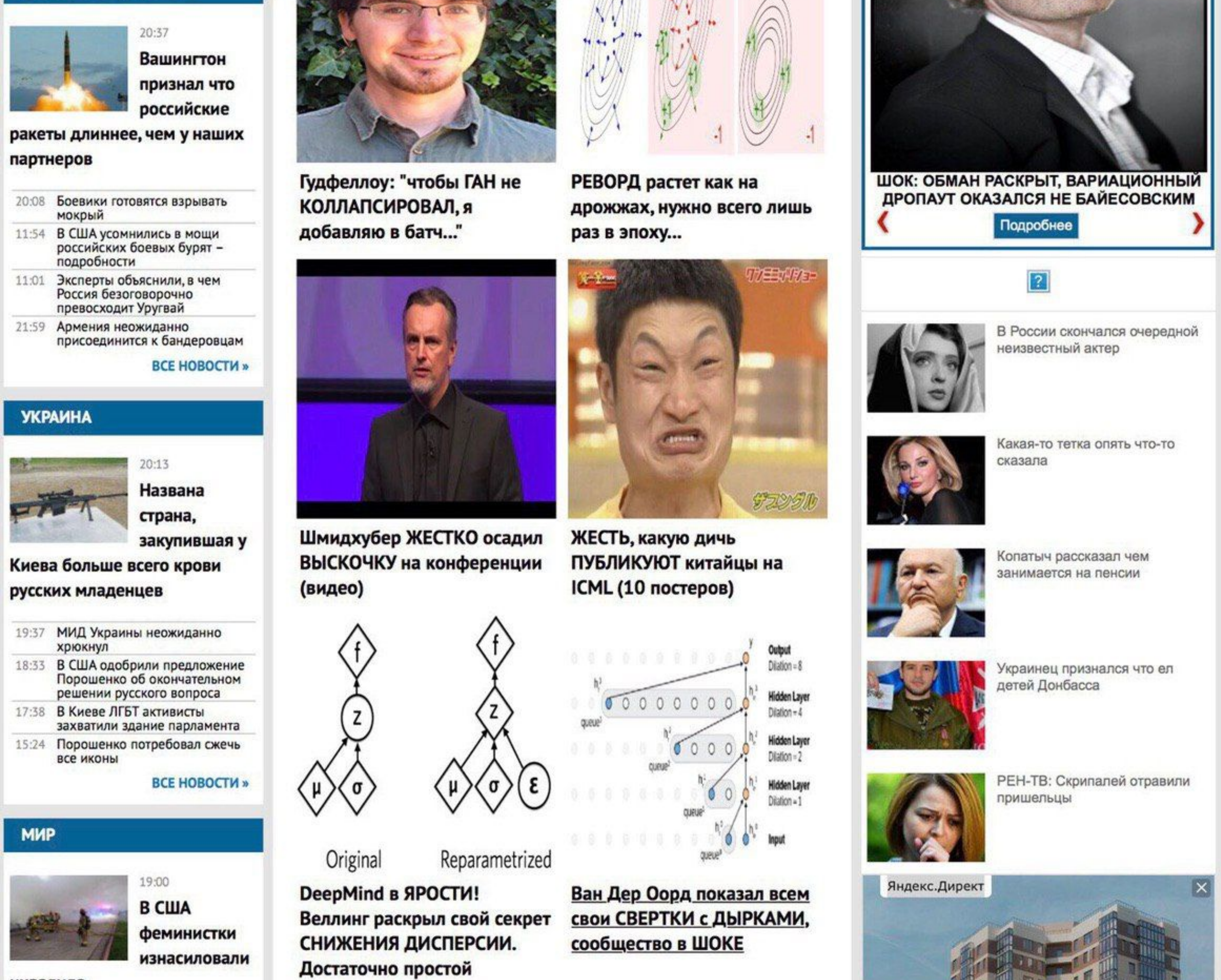
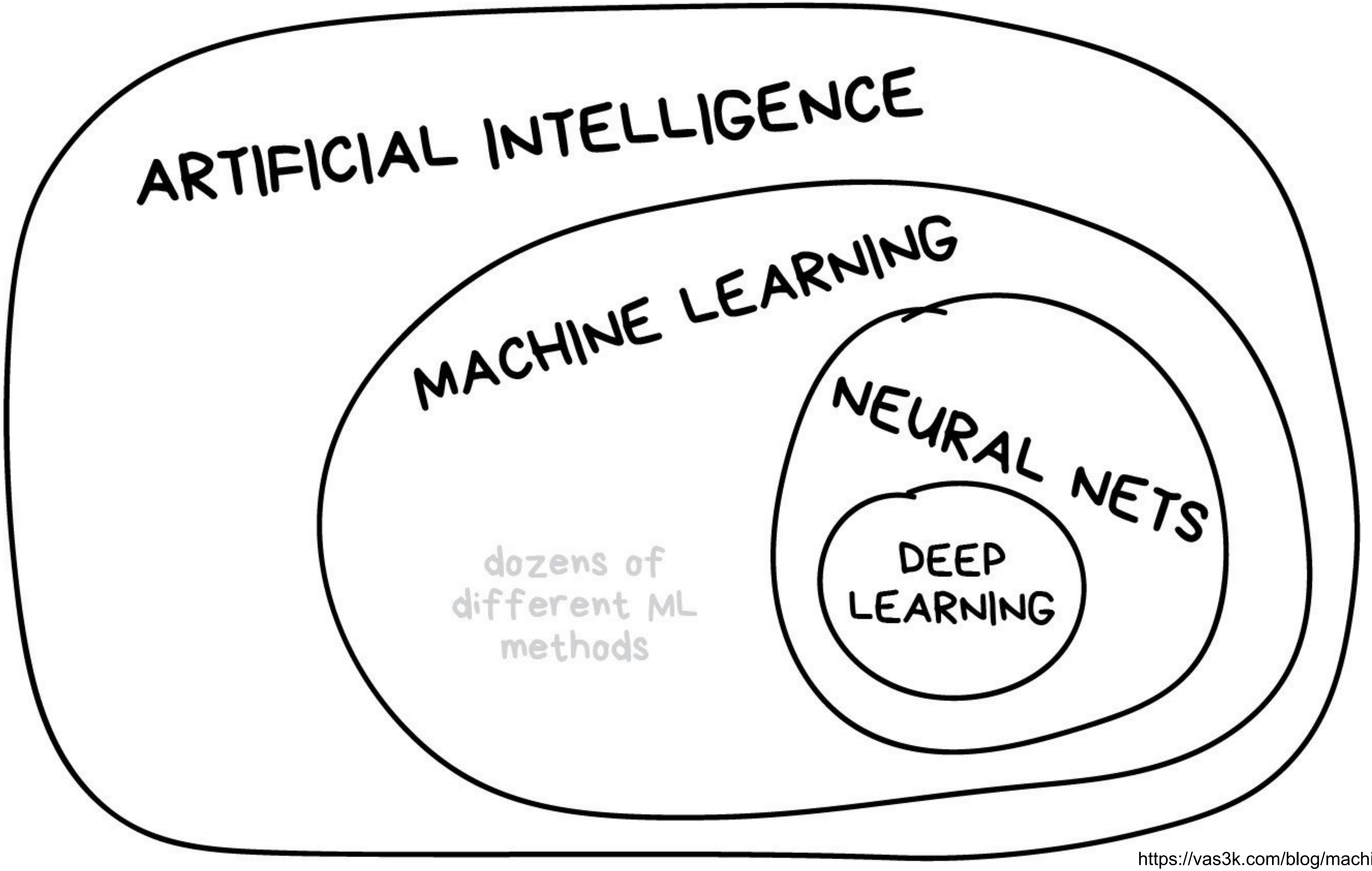


Classical Machine Learning Algorithms

Quick overview

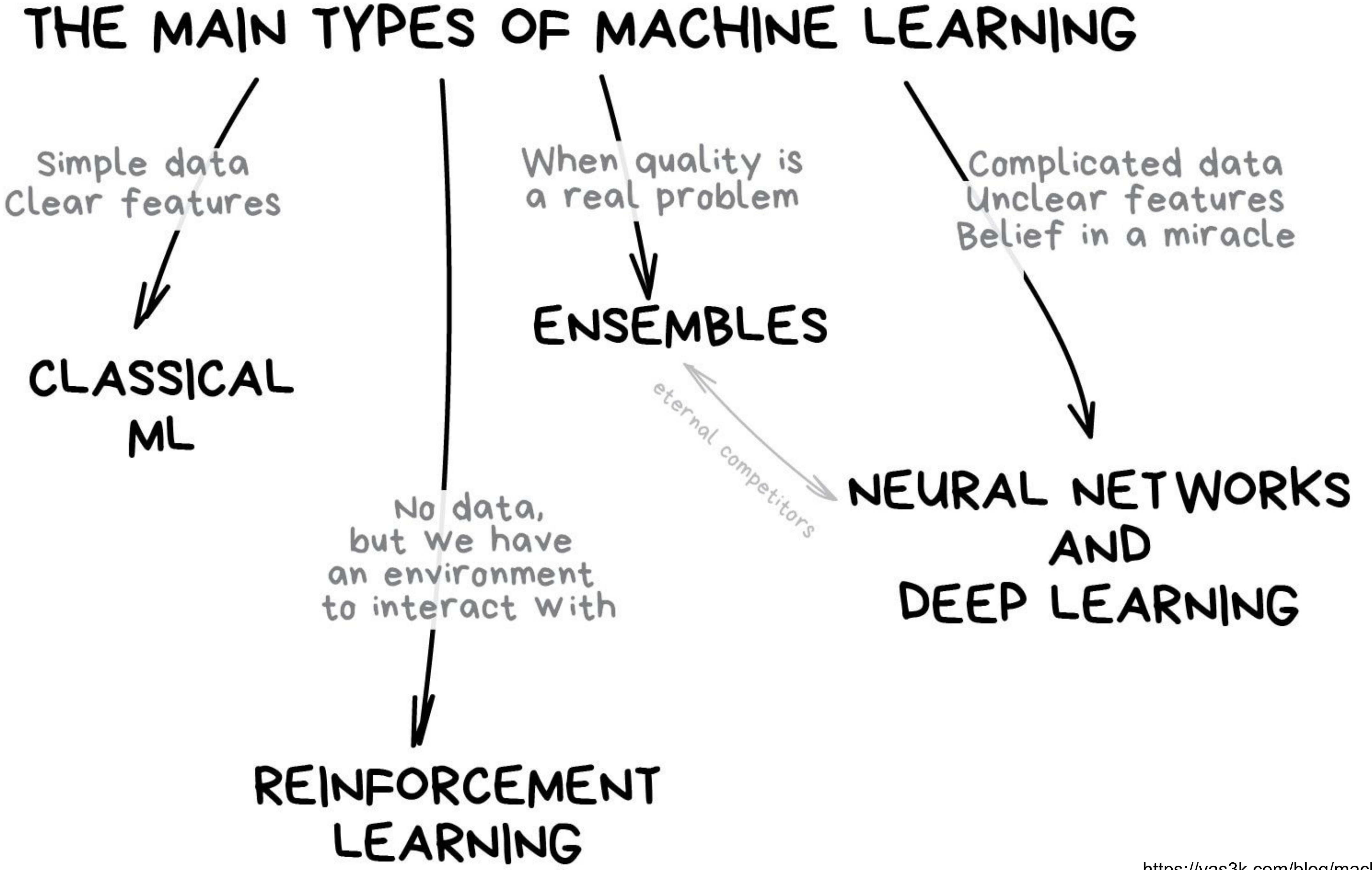
t.me/weirdreparametrizationtrick



https://vas3k.com/blog/machine_learning/

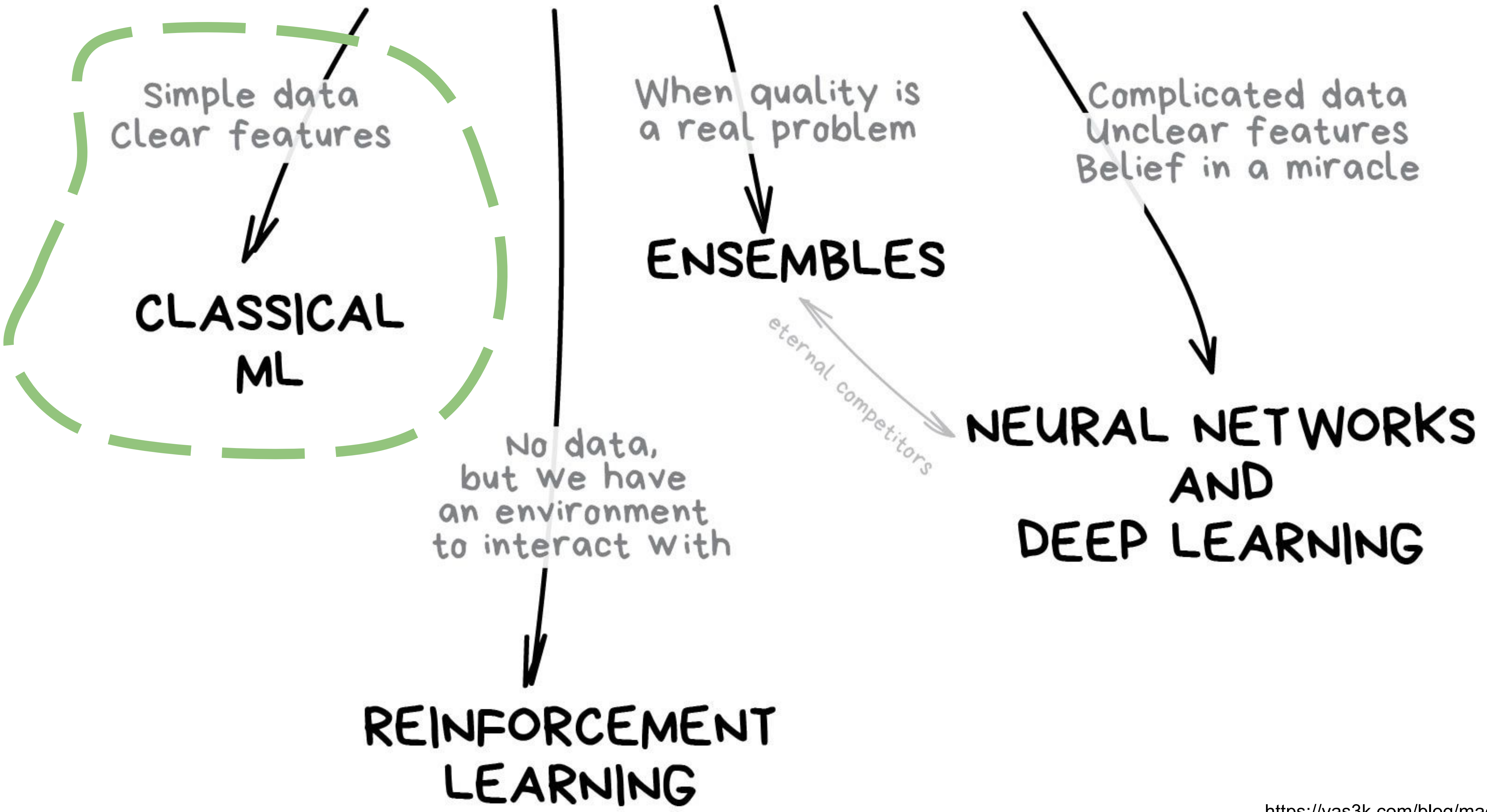
ML Paradigm

E	T	P
Experience	**Task**	**Performance measure**
Relevant data	Expected type of prediction	Metrics for prediction evaluation
Tabular data, labeled images, time series etc.	*Value, class, mask, representation*	*MSE, accuracy, IoU and many more*

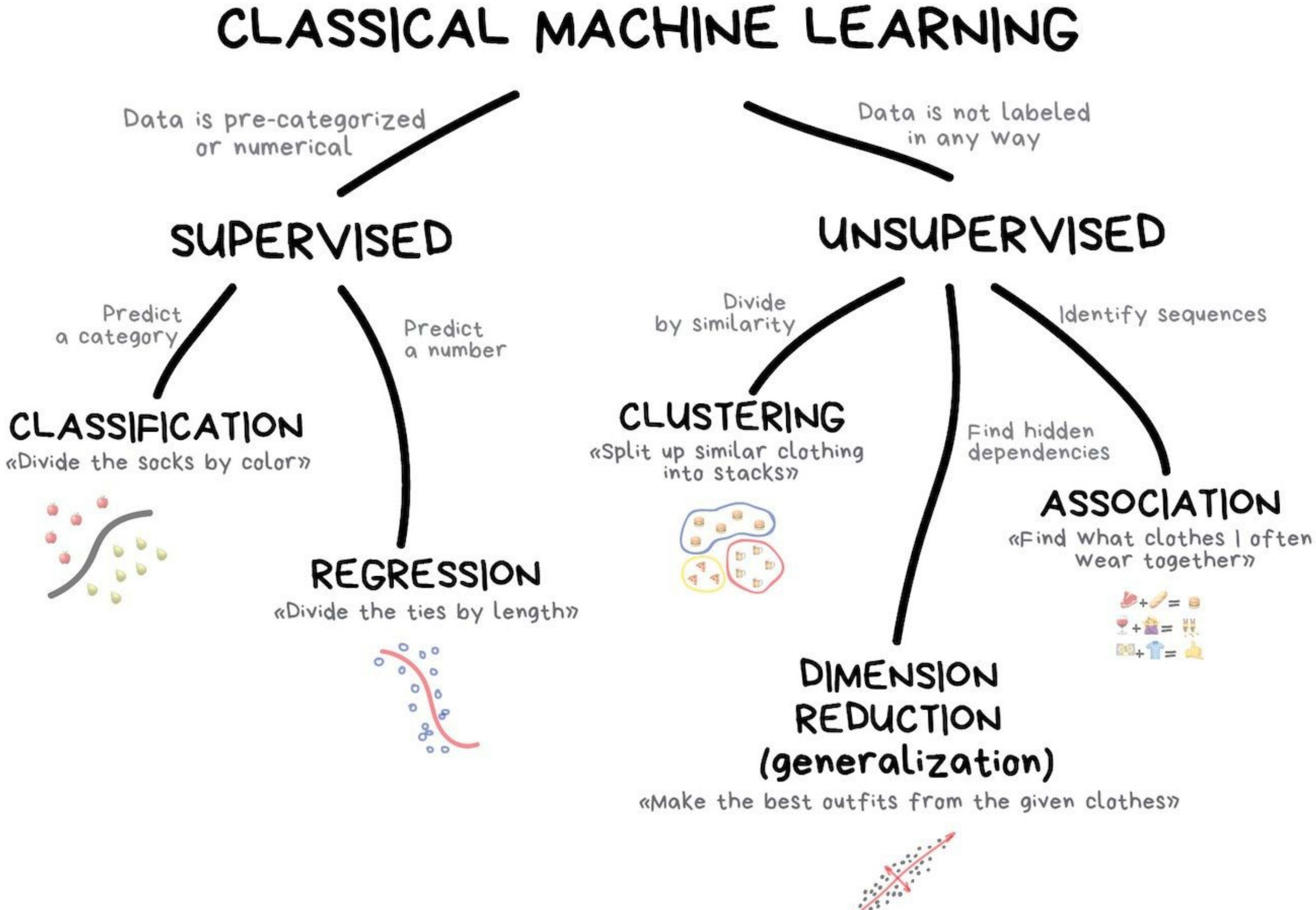


https://vas3k.com/blog/machine_learning/

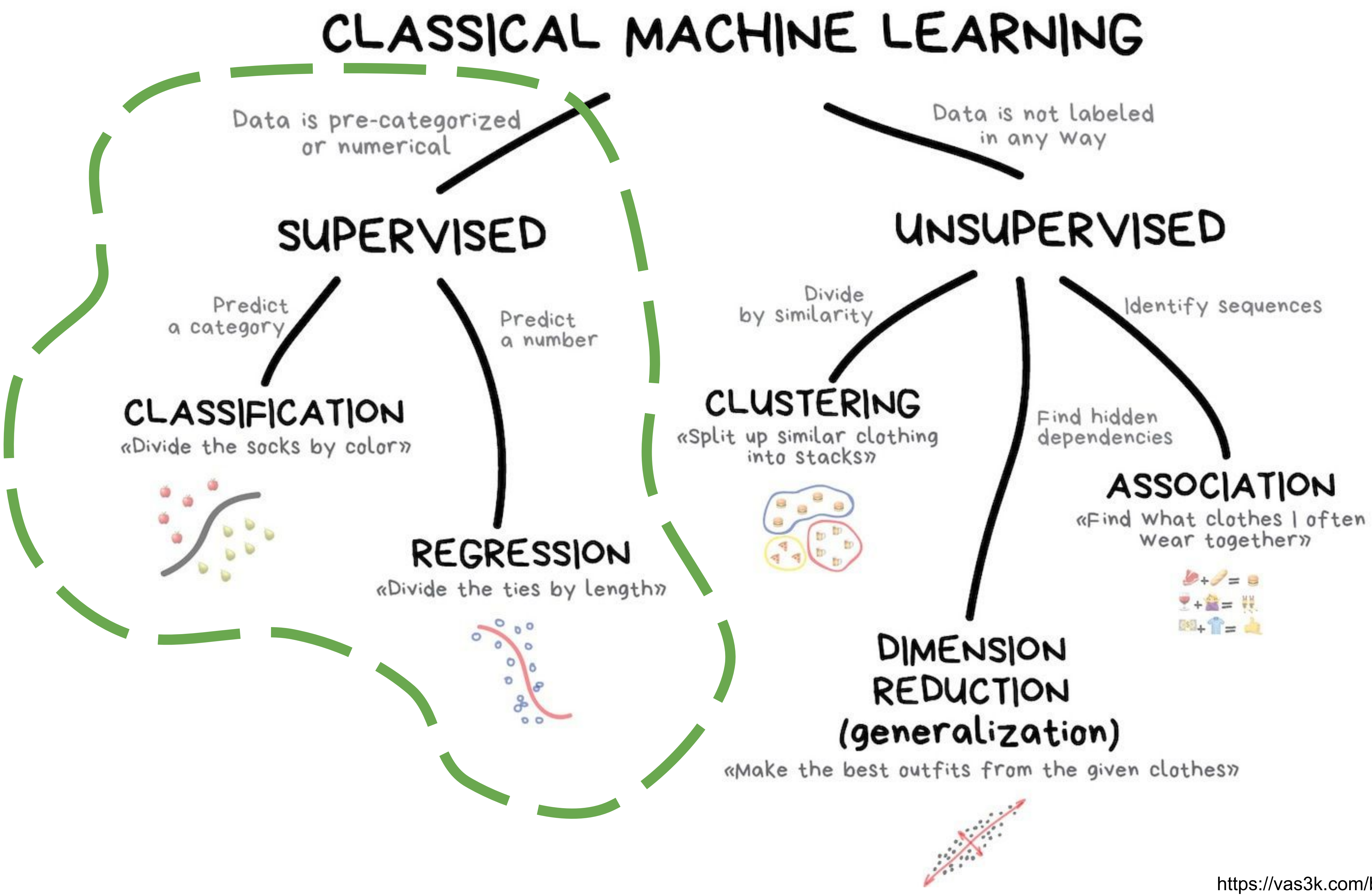
THE MAIN TYPES OF MACHINE LEARNING



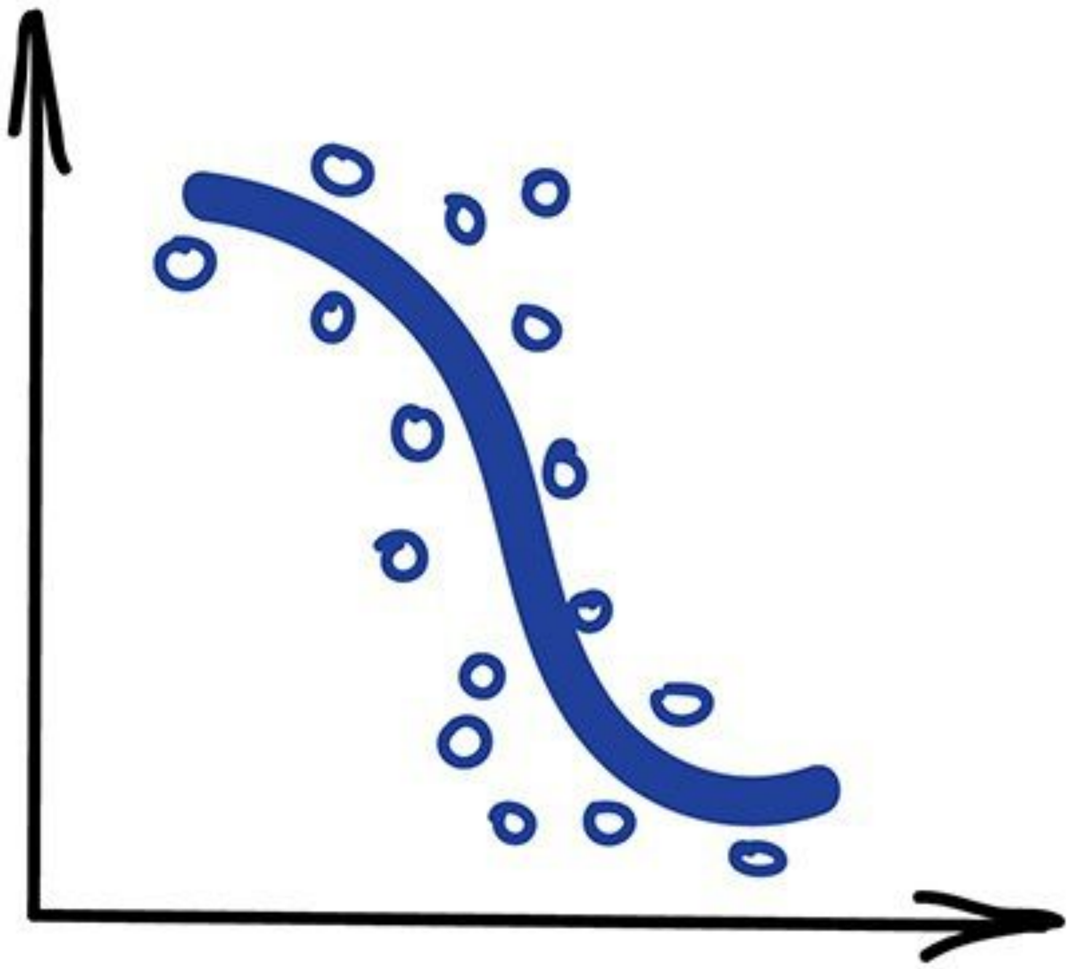
https://vas3k.com/blog/machine_learning/



https://vas3k.com/blog/machine_learning/



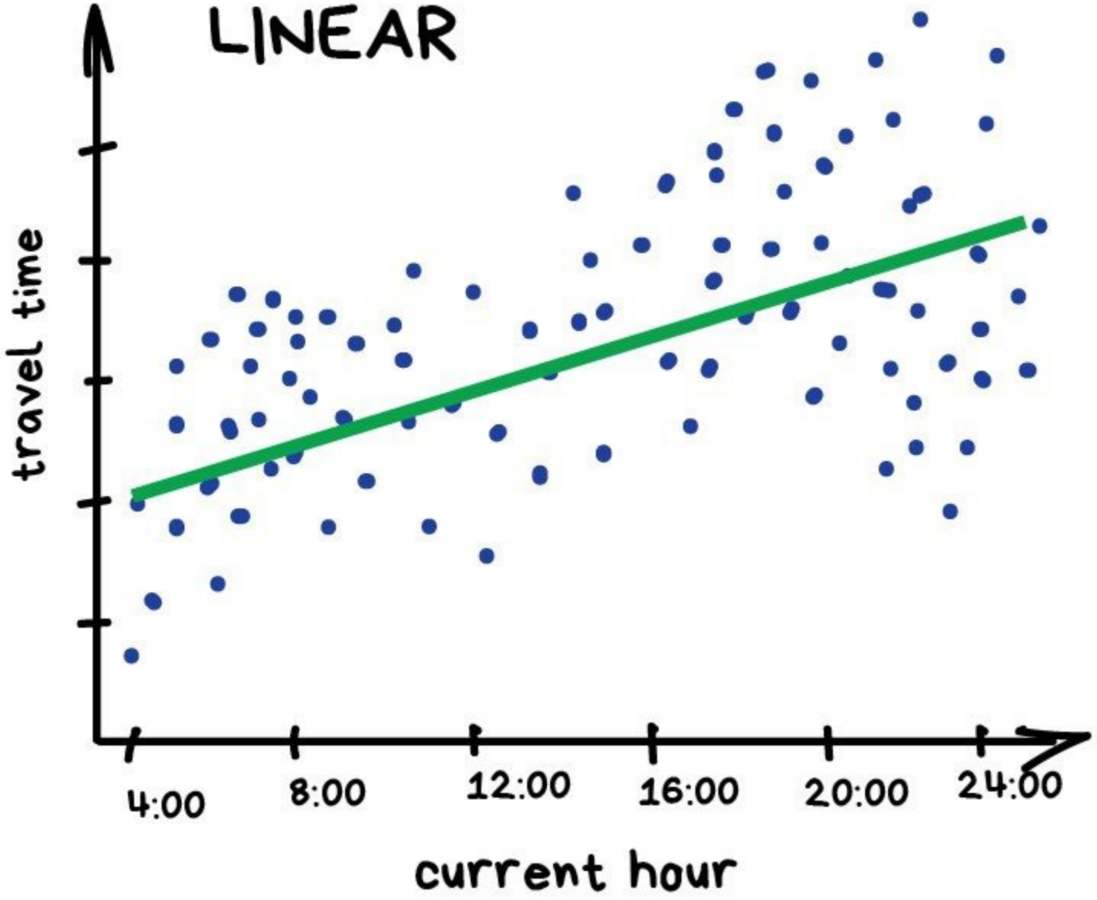
https://vas3k.com/blog/machine_learning/

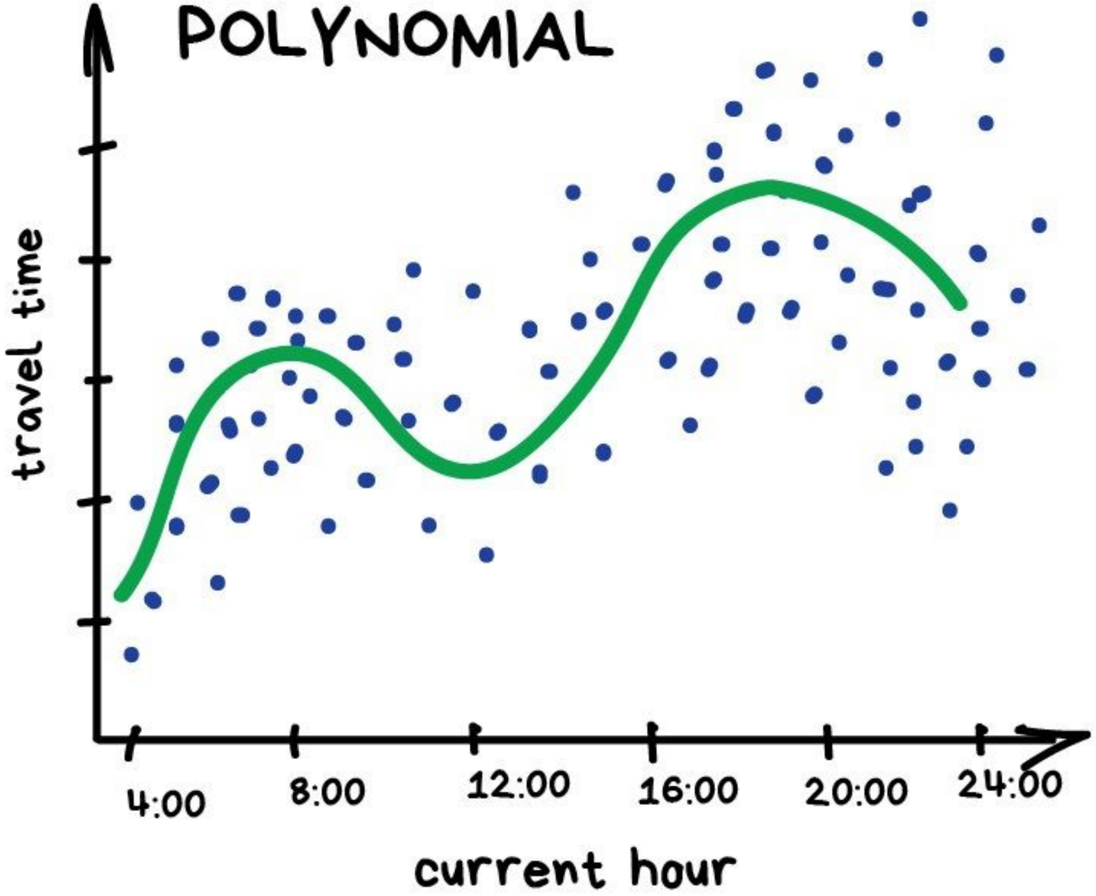


Regression

https://vas3k.com/blog/machine_learning/

PREDICT TRAFFIC JAMS





REGRESSION

https://vas3k.com/blog/machine_learning/

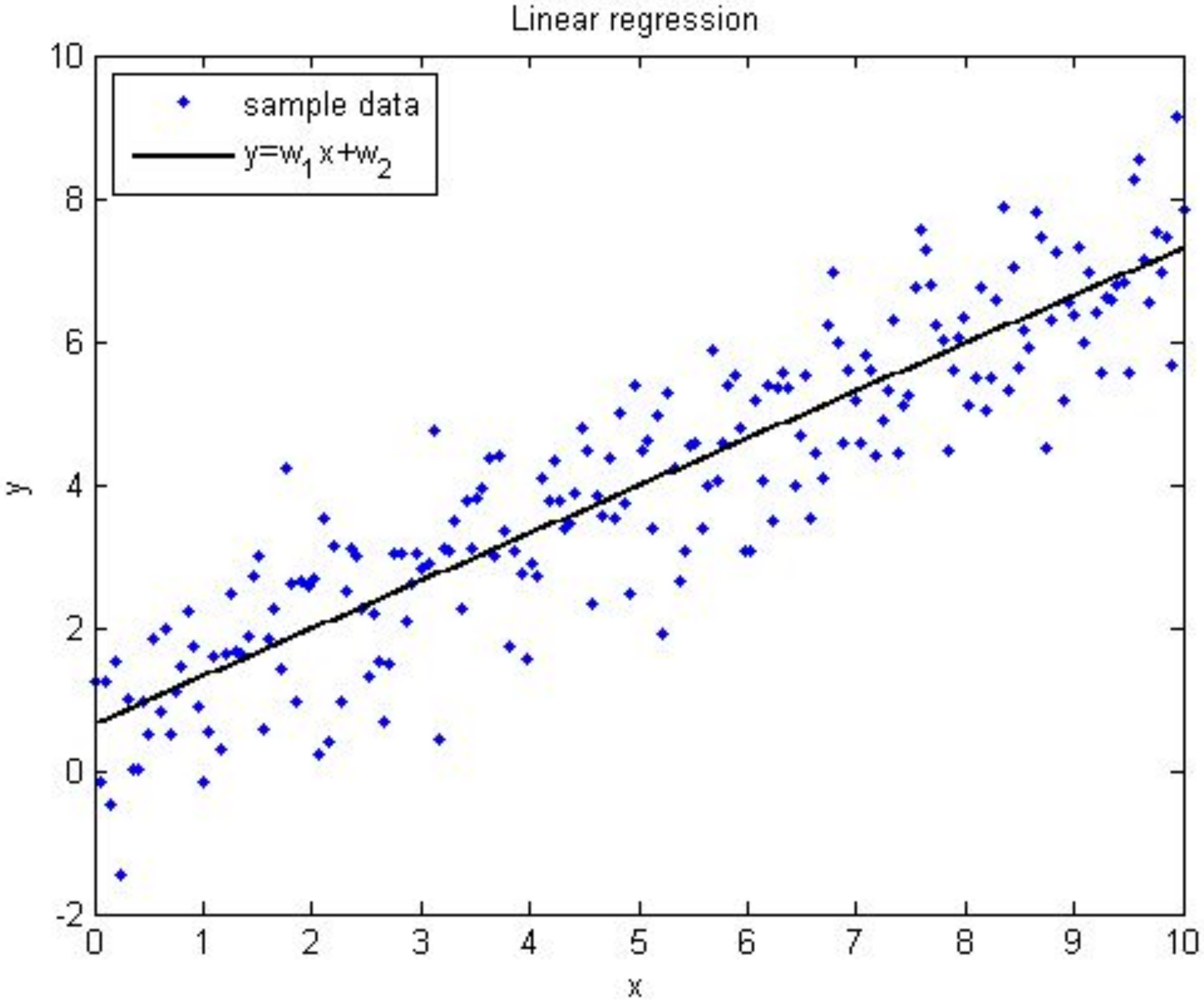
Simple linear regression

$x \in \mathbb{R}$ – input

$y \in \mathbb{R}$ – target

$w_0, w_1 \in \mathbb{R}$ – weights

$\hat{y} = w_0 + w_1 x$



Simple linear regression

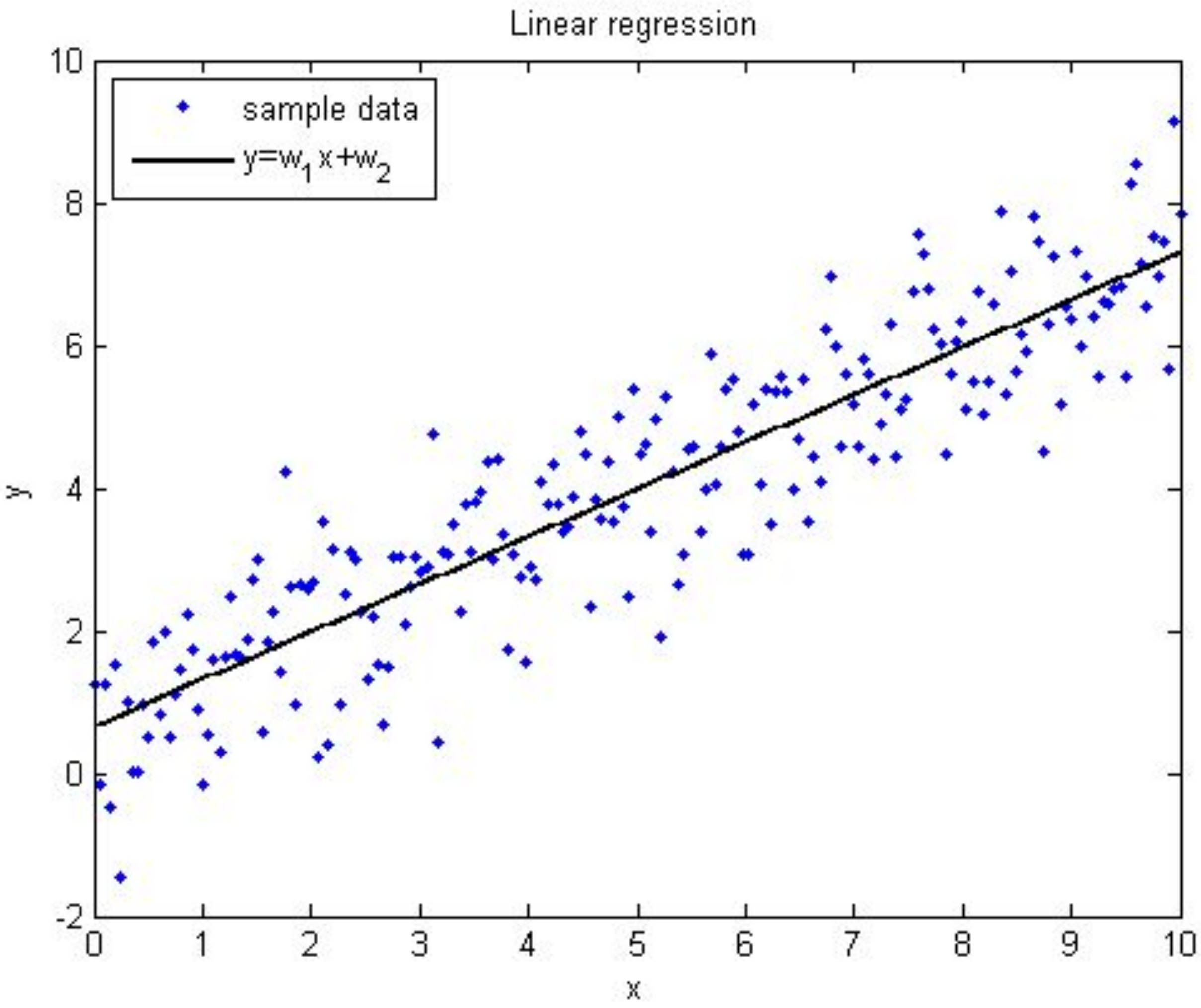
given data

$x \in \mathbb{R}$ – input

$y \in \mathbb{R}$ – target

$w_0, w_1 \in \mathbb{R}$ – weights

$\hat{y} = w_0 + w_1 x$



Simple linear regression

given data

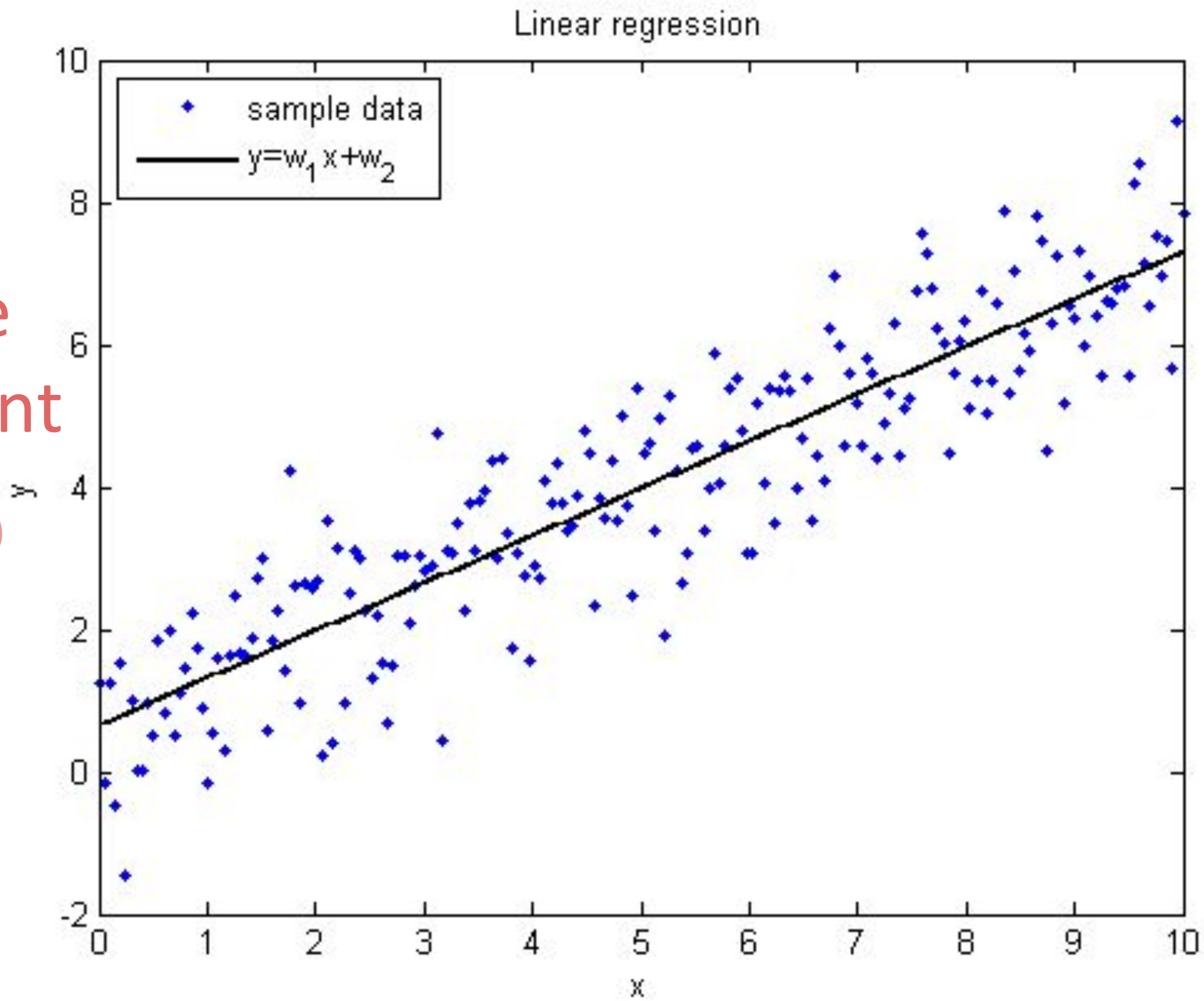
$x \in \mathbb{R}$ – input

$y \in \mathbb{R}$ – target

trainable coefficient

$w_0, w_1 \in \mathbb{R}$ – weights

$$\hat{y} = w_0 + w_1 x$$



Multiple linear regression

housing.csv (1.36 MB) 10 of 10 columns

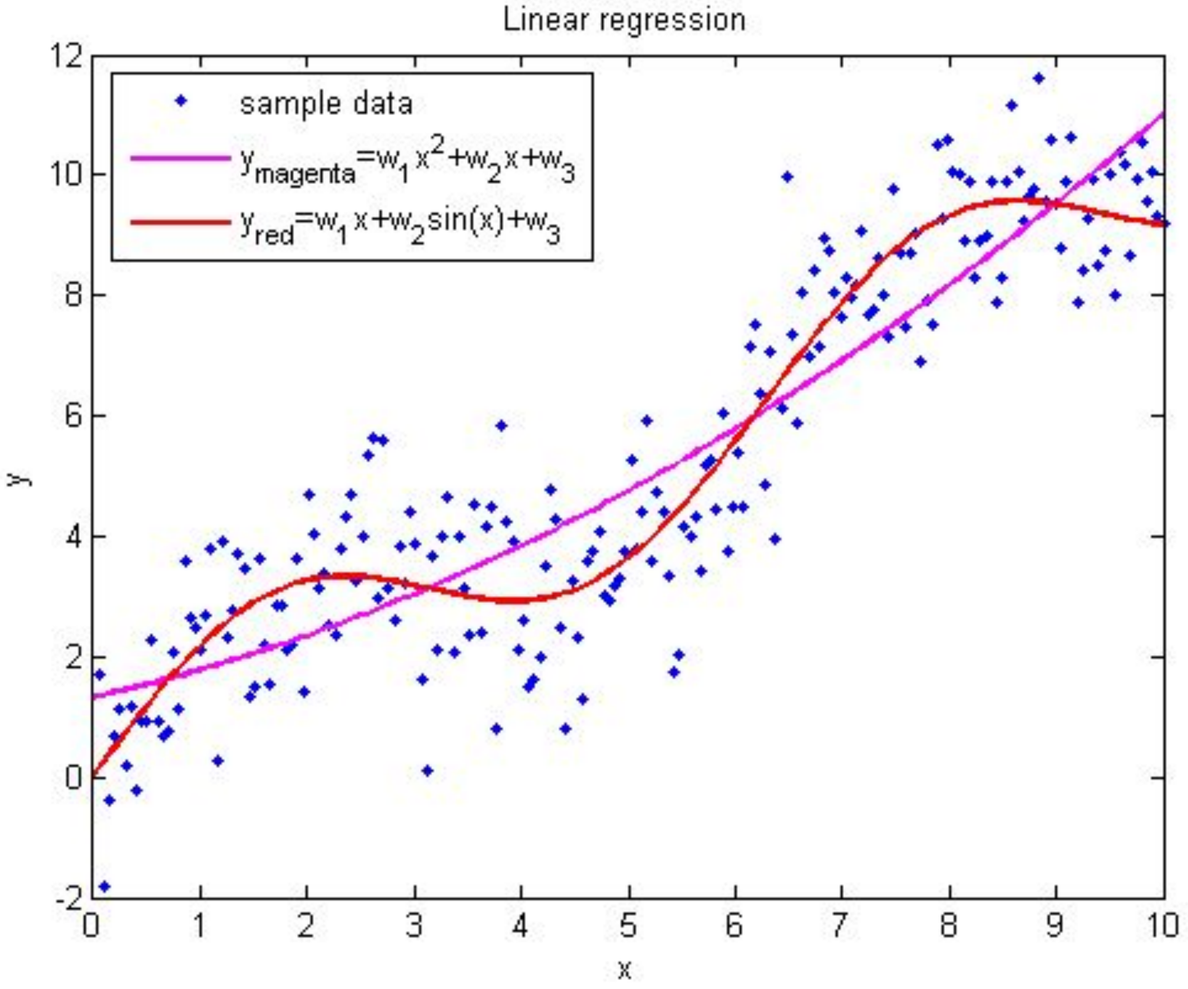
	# longitu...	# latitude	# housin...	# total_ro...	# total_b...	# populat...	# househ...	# median...	# median...	A ocean_...
1	-122.23	37.88	41.0	880.0	129.0	322.0	126.0	8.3252	452600.0	NEAR BAY
2	-122.22	37.86	21.0	7099.0	1106.0	2401.0	1138.0	8.3014	358500.0	NEAR BAY
3	-122.24	37.85	52.0	1467.0	190.0	496.0	177.0	7.2574	352100.0	NEAR BAY
4	-122.25	37.85	52.0	1274.0	235.0	558.0	219.0	5.6431	341300.0	NEAR BAY
5	-122.25	37.85	52.0	1627.0	280.0	565.0	259.0	3.8462	342200.0	NEAR BAY
6	-122.25	37.85	52.0	919.0	213.0	413.0	193.0	4.0368	269700.0	NEAR BAY
7	-122.25	37.84	52.0	2535.0	489.0	1094.0	514.0	3.6591	299200.0	NEAR BAY
8	-122.25	37.84	52.0	3104.0	687.0	1157.0	647.0	3.12	241400.0	NEAR BAY
9	-122.26	37.84	42.0	2555.0	665.0	1206.0	595.0	2.0804	226700.0	NEAR BAY
10										AY

$$\hat{y} = w_0 + w_1 x_1 + \ldots + w_N x_N = \sum_{j=0}^{M} w_j x_j$$

Also linear regression

$$\hat{y} = \sum_{j=0}^{M} w_j g(x_j)$$

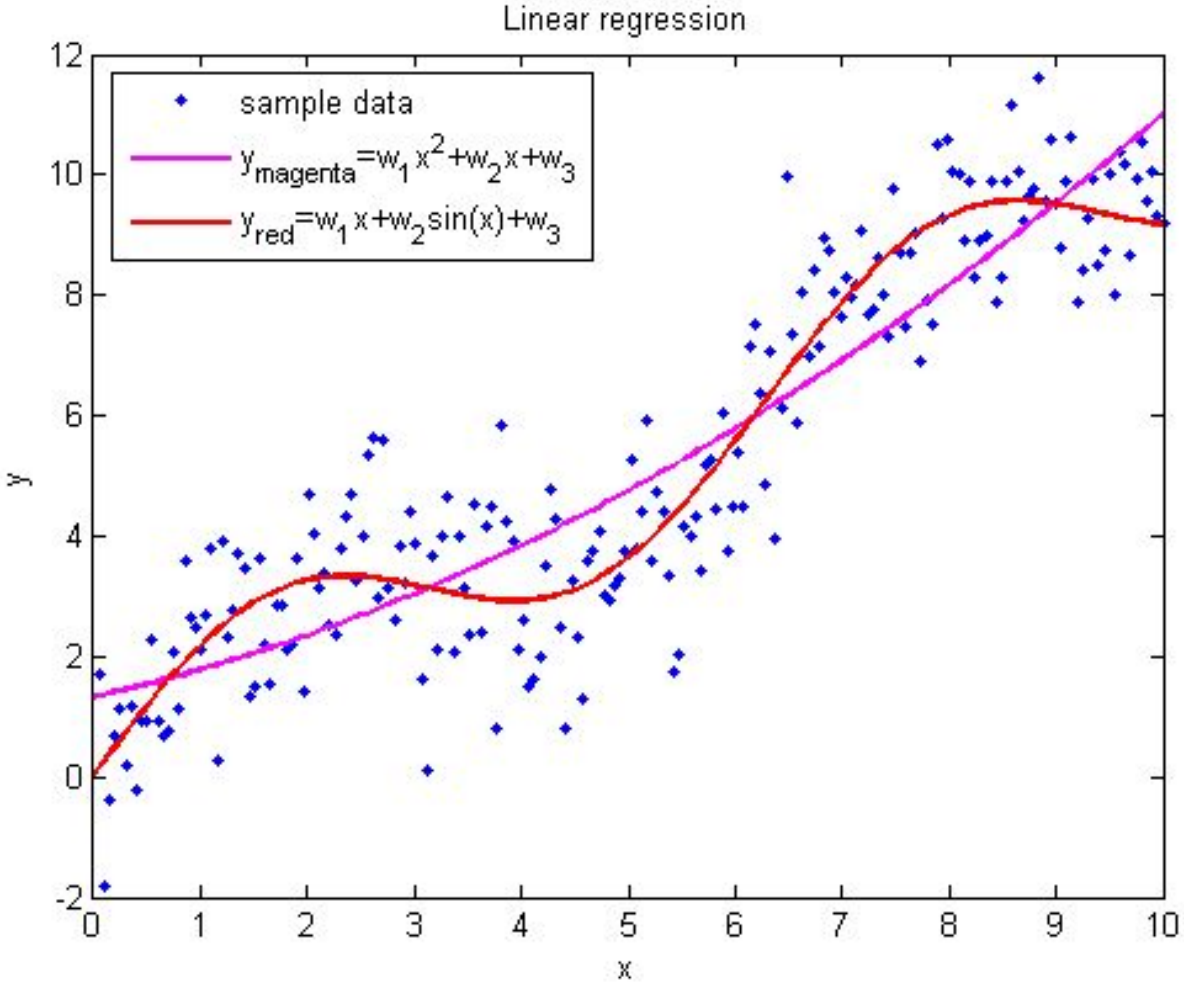
$$g(x) = x \Rightarrow \hat{y} = \sum_{j=0}^{M} w_j x_j$$



Also linear regression

$$\hat{y} = \sum_{j=0}^{M} w_j g(x_j)$$

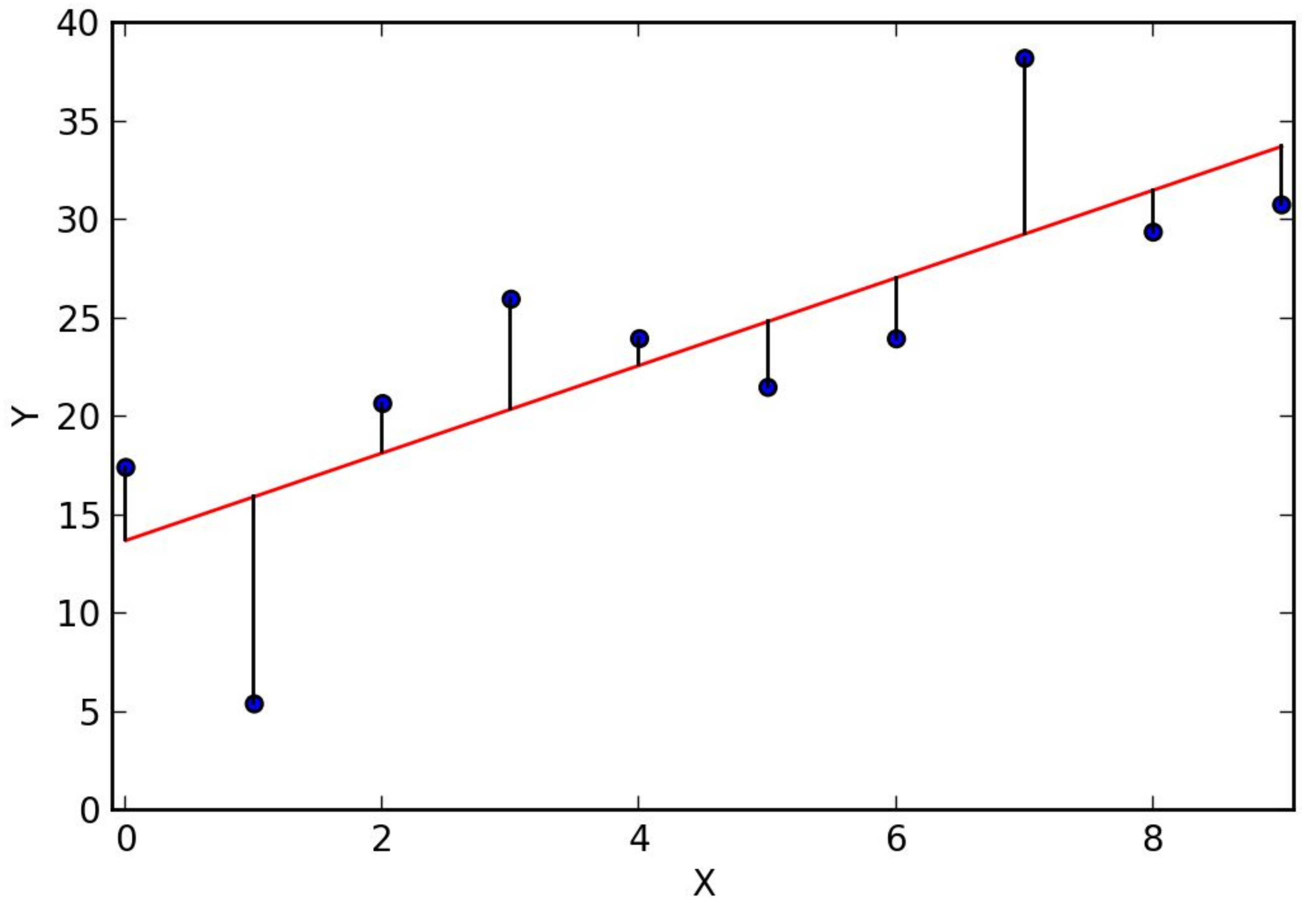
$$g(x) = x \Rightarrow \hat{y} = \sum_{j=0}^{M} w_j x_j$$



Evaluating prediction

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

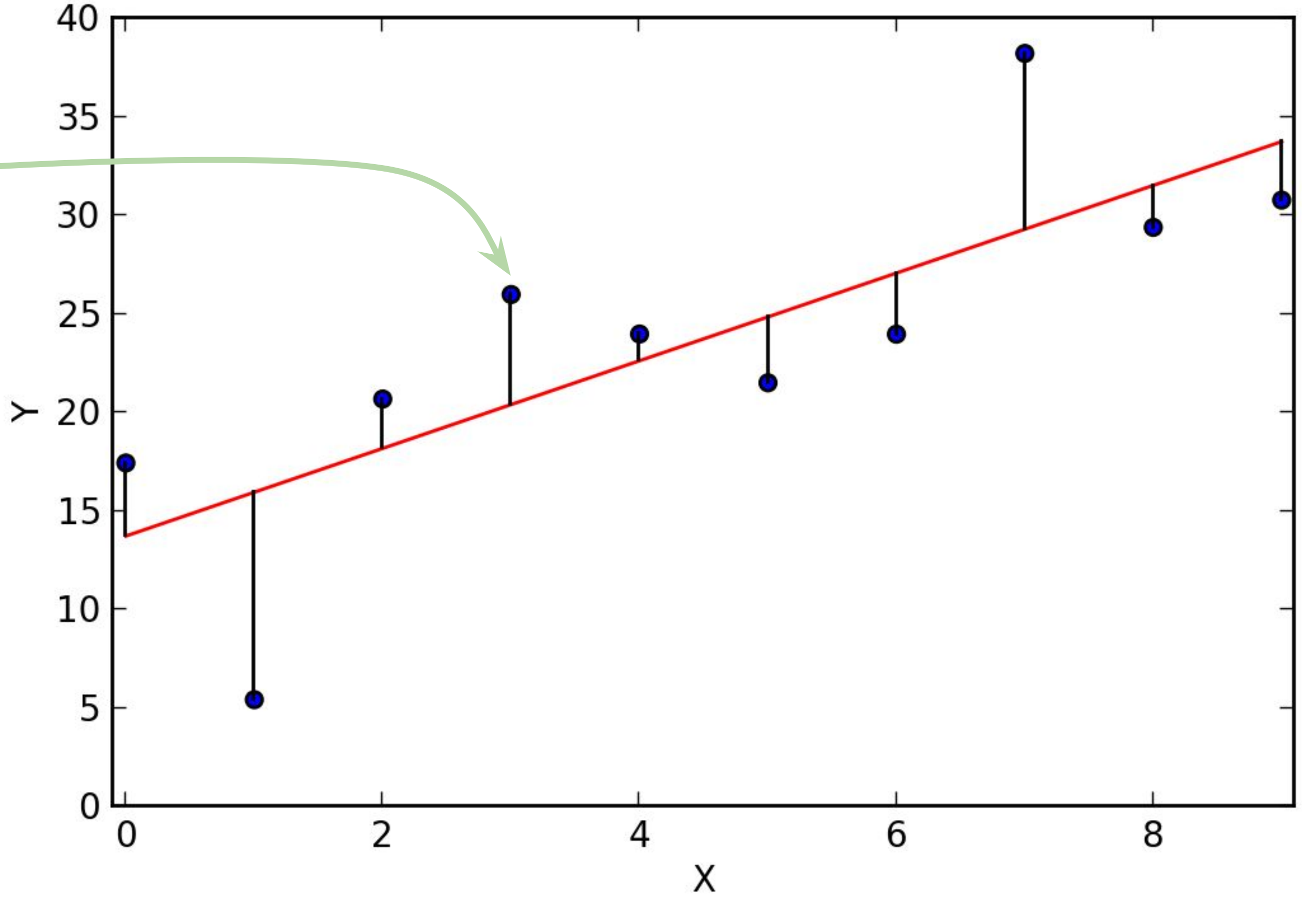
y – target
$\hat{y}$ – prediction



Evaluating prediction

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

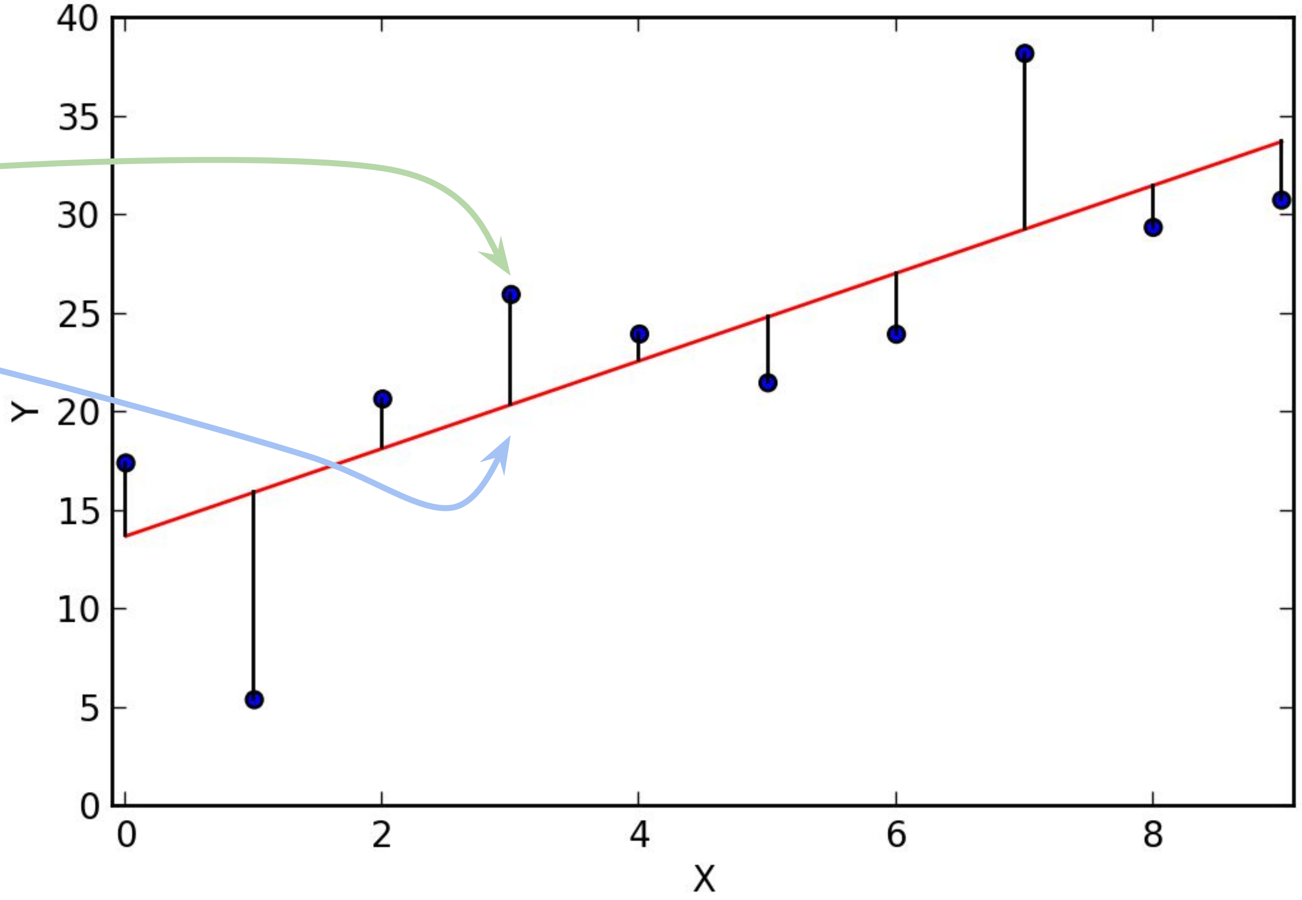
y – target
$\hat{y}$ – prediction



Evaluating prediction

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

y – target
$\hat{y}$ – prediction



Training parameters

$$\hat{y} = w_0 + w_1 x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$
$$= (y - w_1 x - w_0)^2$$

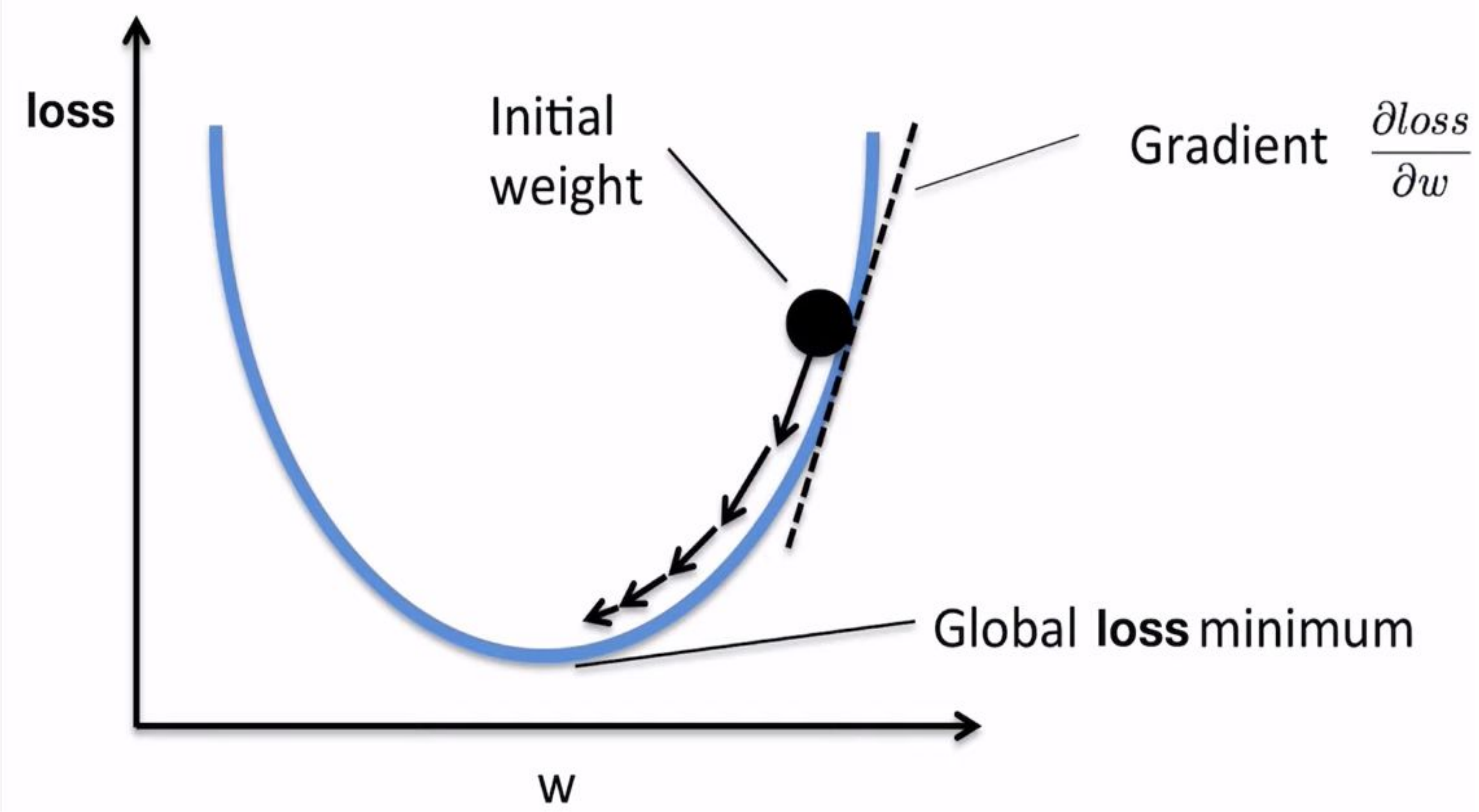
Training parameters

$$\hat{y} = w_0 + w_1 x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$
$$= (y - w_1 x - w_0)^2$$

$$w_{new} = w_{old} - \alpha \frac{\partial L(w,x)}{\partial w}$$
$\alpha \in \mathbb{R}$ – learning rate

Training parameters



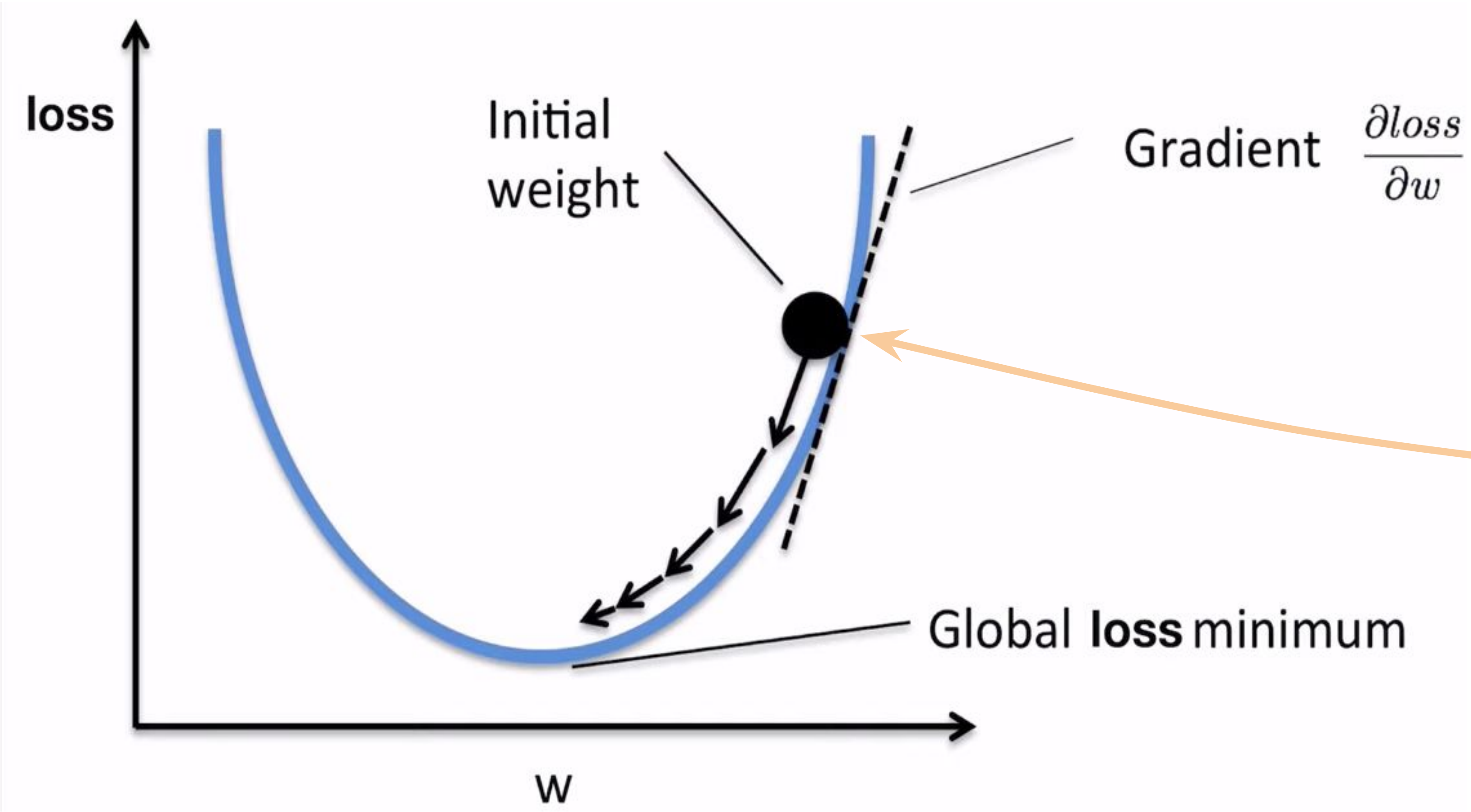
$$\hat{y} = w_0 + w_1 x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$
$$= (y - w_1 x - w_0)^2$$

$$w_{new} = w_{old} - \alpha \frac{\partial L(w,x)}{\partial w}$$

$\alpha \in \mathbb{R}$ – learning rate

Training parameters



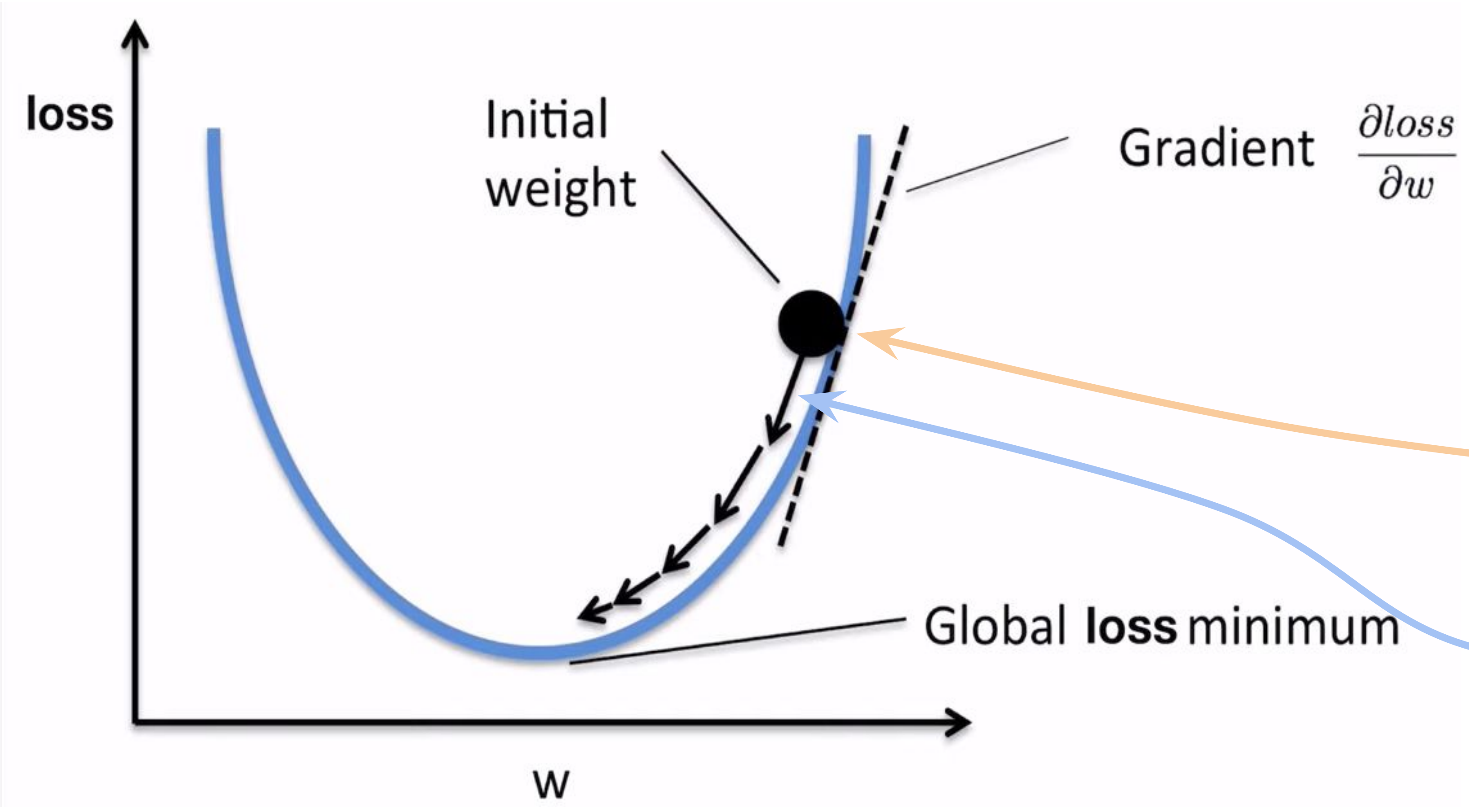
$$\hat{y} = w_0 + w_1 x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$
$$= (y - w_1 x - w_0)^2$$

$$w_{new} = w_{old} - \alpha \frac{\partial L(w,x)}{\partial w}$$

$\alpha \in \mathbb{R}$ – learning rate

Training parameters



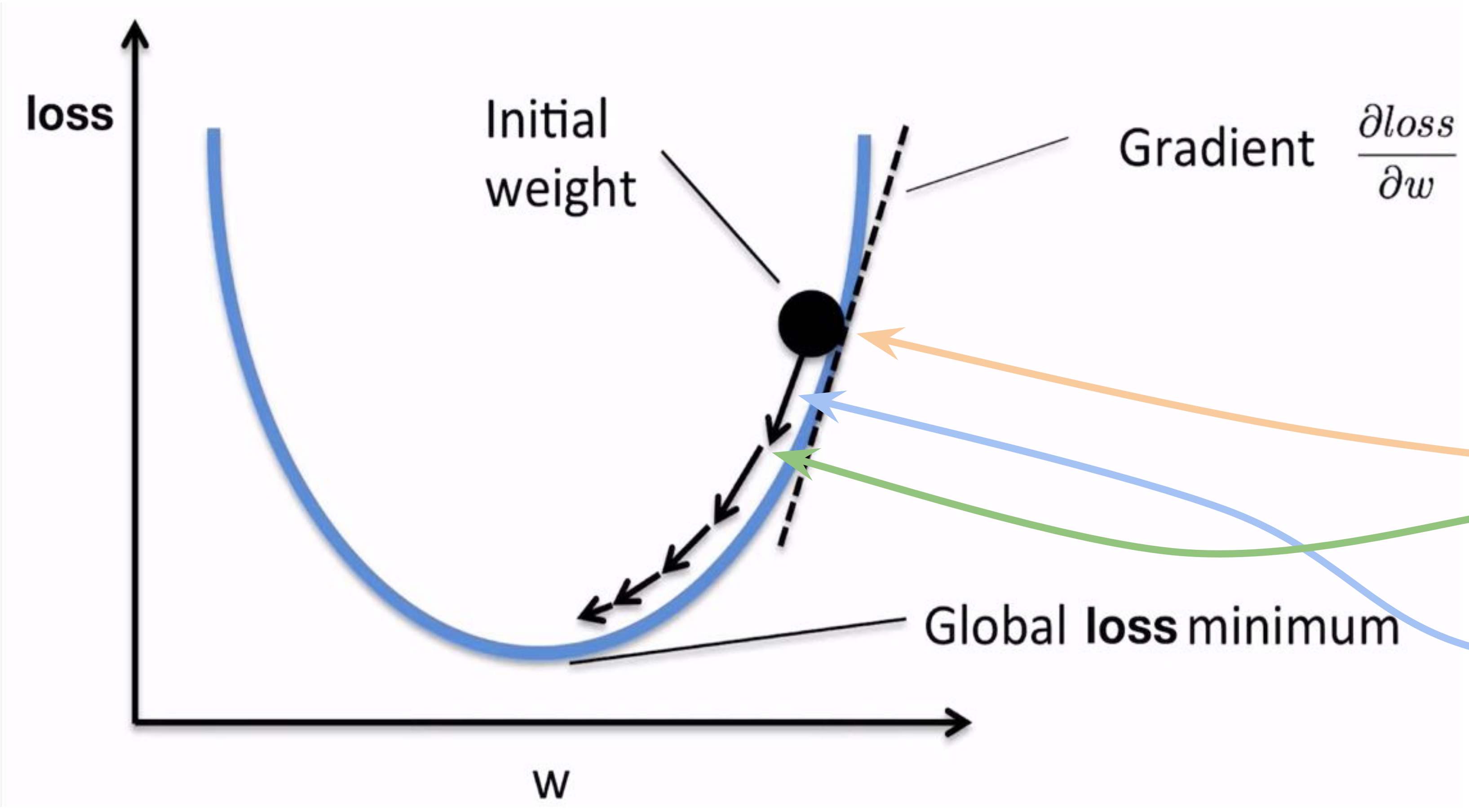
$$\hat{y} = w_0 + w_1 x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$
$$= (y - w_1 x - w_0)^2$$

$$w_{new} = w_{old} - \alpha \frac{\partial L(w,x)}{\partial w}$$

$\alpha \in \mathbb{R}$ – learning rate

Training parameters



$$\hat{y} = w_0 + w_1 x$$

$$L(y, \hat{y}) = (y - \hat{y})^2$$
$$= (y - w_1 x - w_0)^2$$

$$w_{new} = w_{old} - \alpha \frac{\partial L(w,x)}{\partial w}$$

$\alpha \in \mathbb{R}$ – learning rate

Linear regression overview

special case: $\hat{y} = w_1 x + w_0$

Linear regression overview

special case: $\hat{y} = w_1 x + w_0$

general case: $\hat{y} = w_0 + w_1 x_1 + \ldots + w_N x_N = \sum_{j=0}^{M} w_j x_j$

Linear regression overview

special case: $\hat{y} = w_1 x + w_0$

general case: $\hat{y} = w_0 + w_1 x_1 + \ldots + w_N x_N = \sum_{j=0}^{M} w_j x_j$

even more general case: $\hat{y} = \sum_{j=0}^{M} w_j g(x_j)$

Linear regression overview

special case: $\hat{y} = w_1 x + w_0$

general case: $\hat{y} = w_0 + w_1 x_1 + \ldots + w_N x_N = \sum_{j=0}^{M} w_j x_j$

even more general case: $\hat{y} = \sum_{j=0}^{M} w_j g(x_j)$

loss function: $\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

Linear regression overview

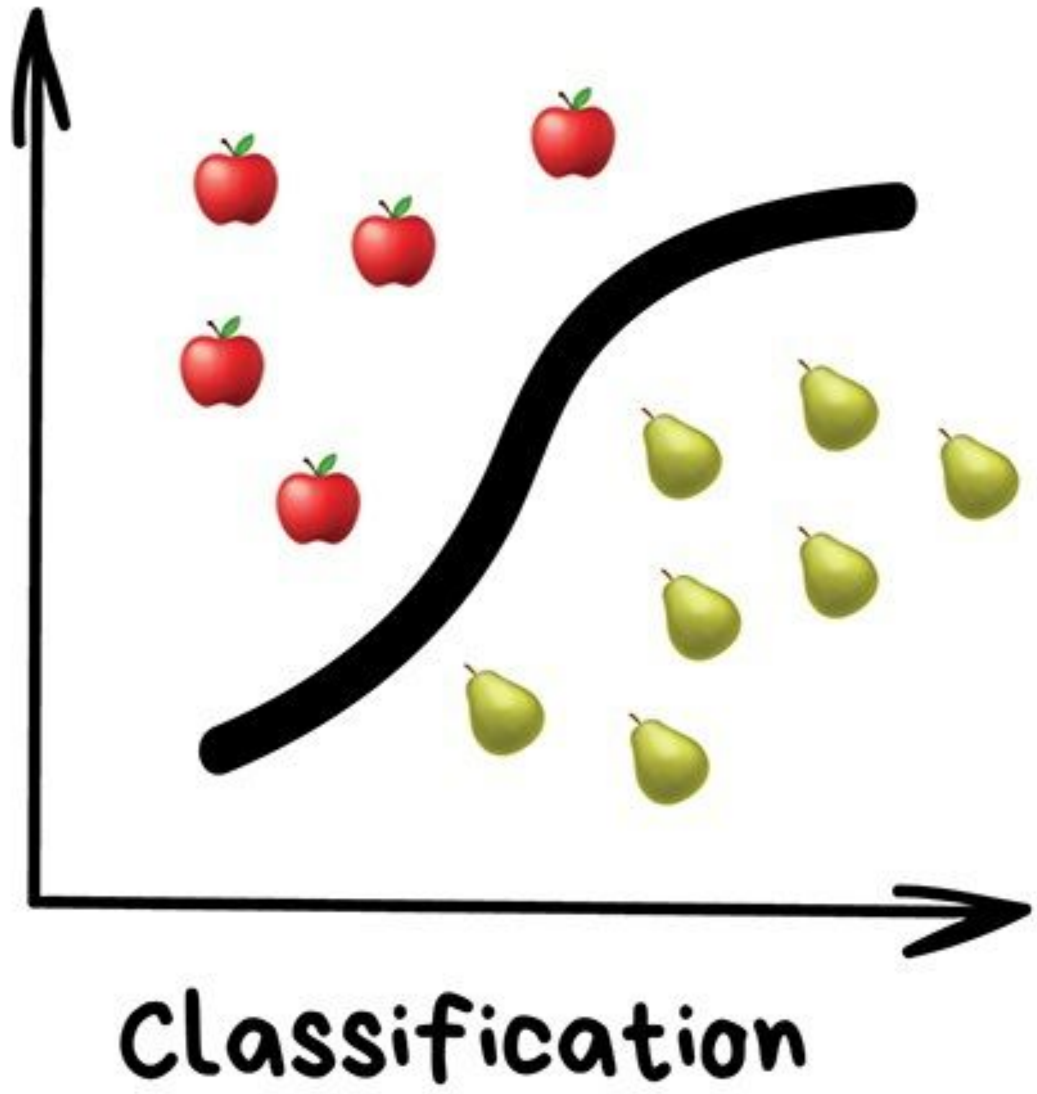
special case: $\hat{y} = w_1 x + w_0$

general case: $\hat{y} = w_0 + w_1 x_1 + \ldots + w_N x_N = \sum_{j=0}^{M} w_j x_j$

even more general case: $\hat{y} = \sum_{j=0}^{M} w_j g(x_j)$

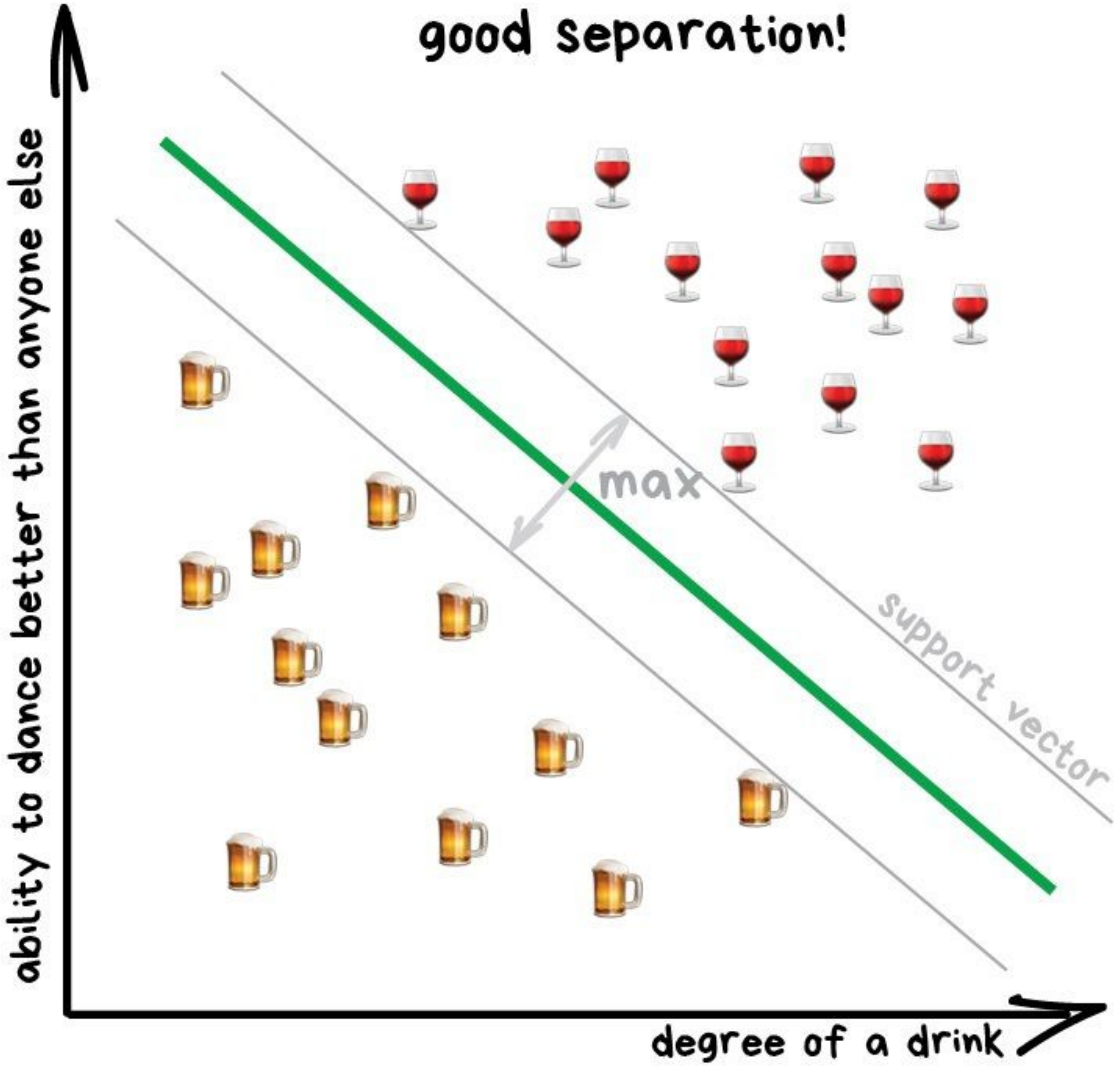
loss function: $\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$

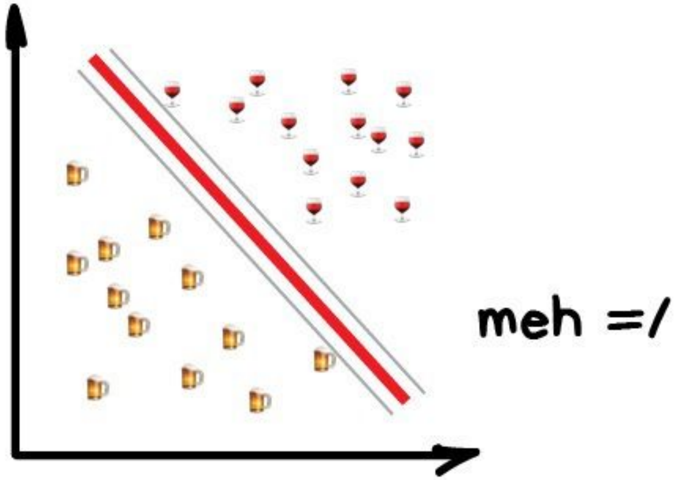
weights update: $w_{new} = w_{old} - \frac{\partial L}{\partial w}$

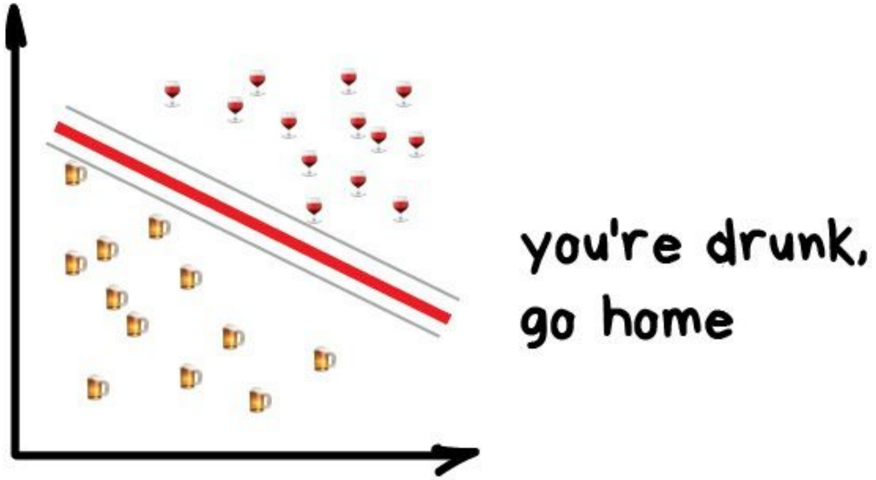


https://vas3k.com/blog/machine_learning/

SEPARATE TYPES OF ALCOHOL







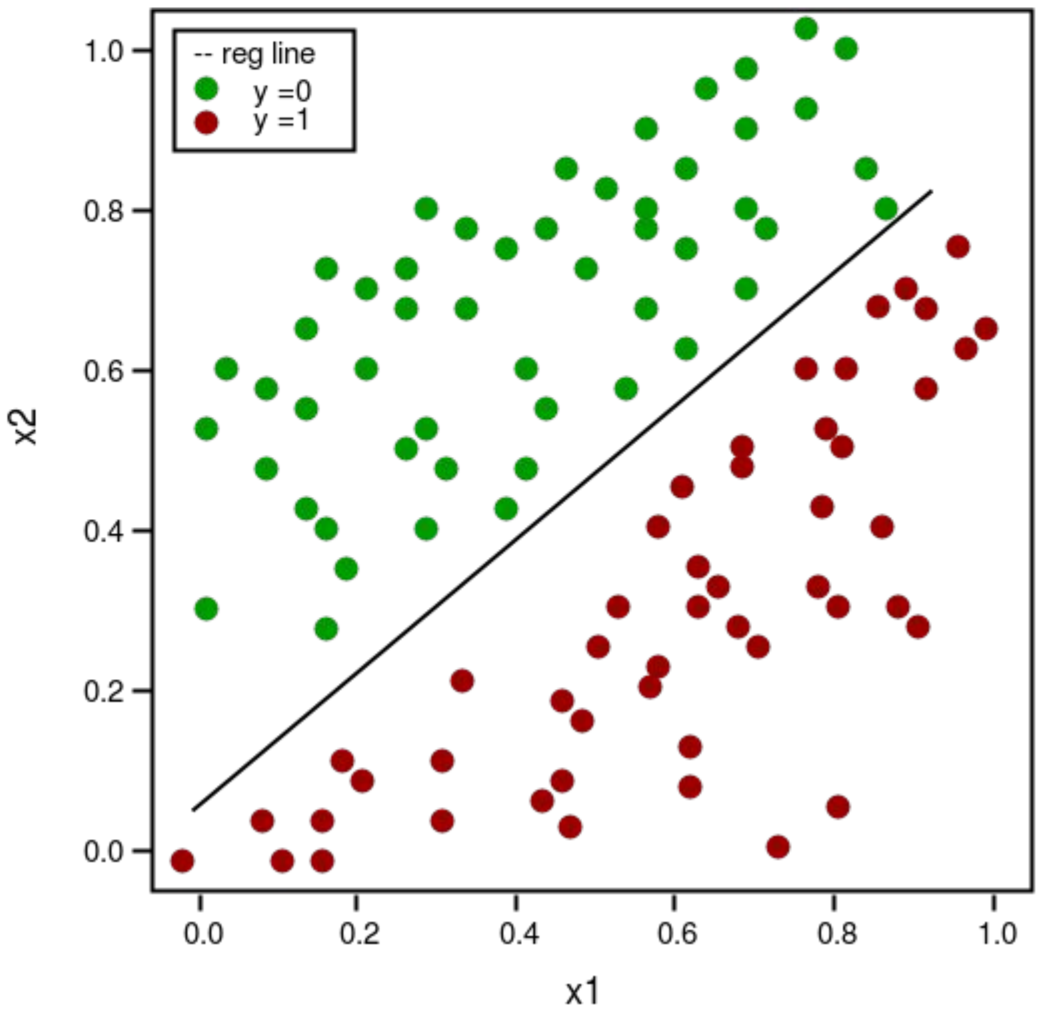
https://vas3k.com/blog/machine_learning/

Logistic regression

$x_1, x_2 \in \mathbb{R}$ – inputs
$y \in \{0, 1\}$ – target

Logistic regression

$x_1, x_2 \in \mathbb{R}$ – inputs
$y \in \{0, 1\}$ – target



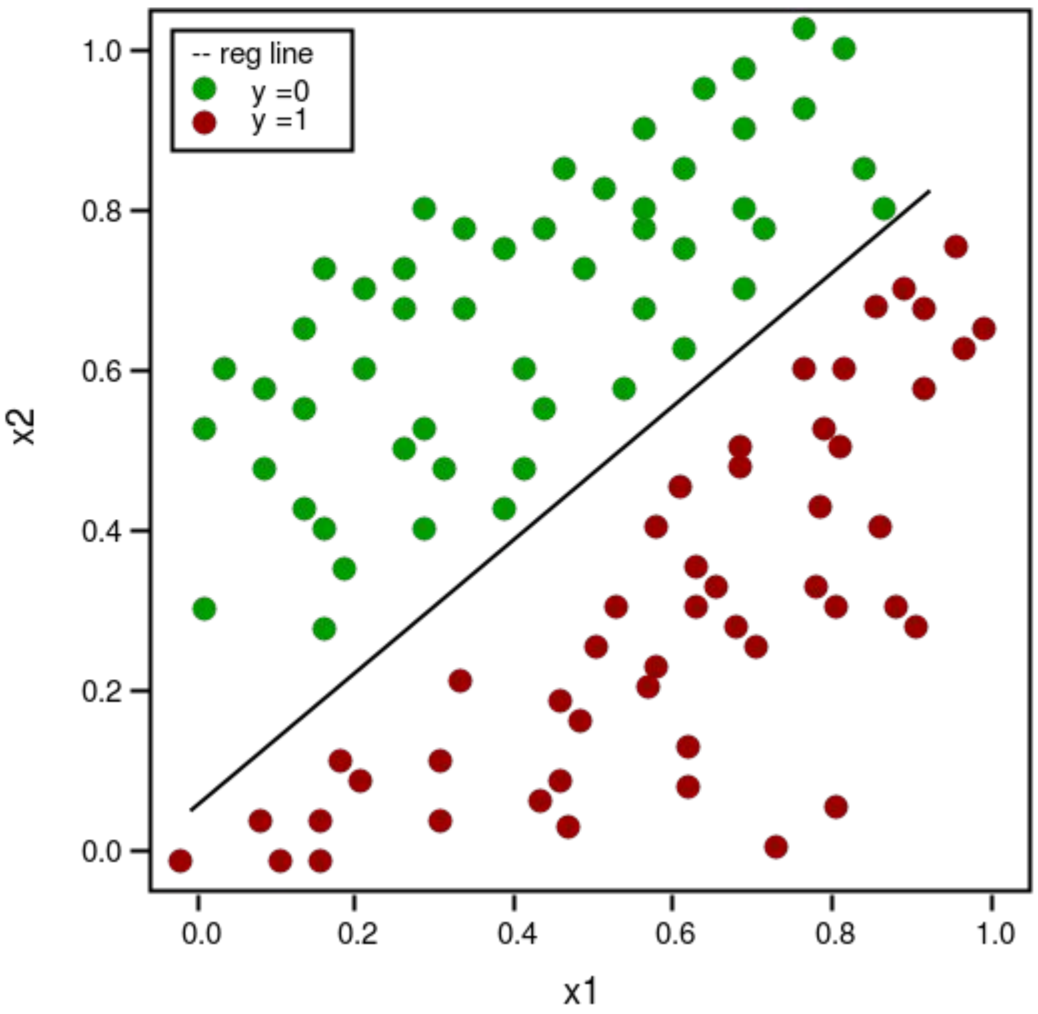
Logistic regression

$x_1, x_2 \in \mathbb{R}$ – inputs

$y \in \{0, 1\}$ – target

$w_0, w_1, w_2 \in \mathbb{R}$ – weights

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2$$



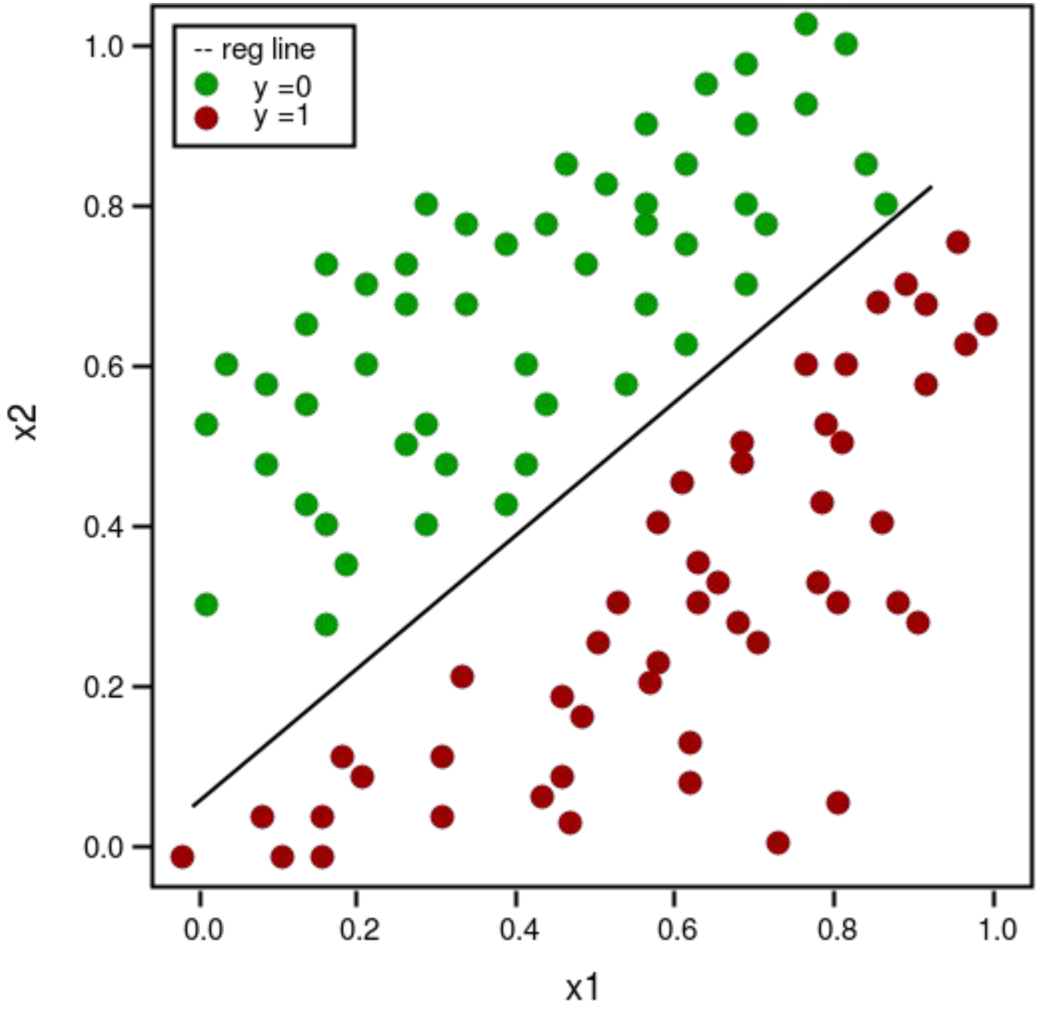
Logistic regression

$x_1, x_2 \in \mathbb{R}$ – inputs

$y \in \{0, 1\}$ – target

$w_0, w_1, w_2 \in \mathbb{R}$ – weights

$\cancel{\hat{y} = w_0 + w_1 x_1 + w_2 x_2}$



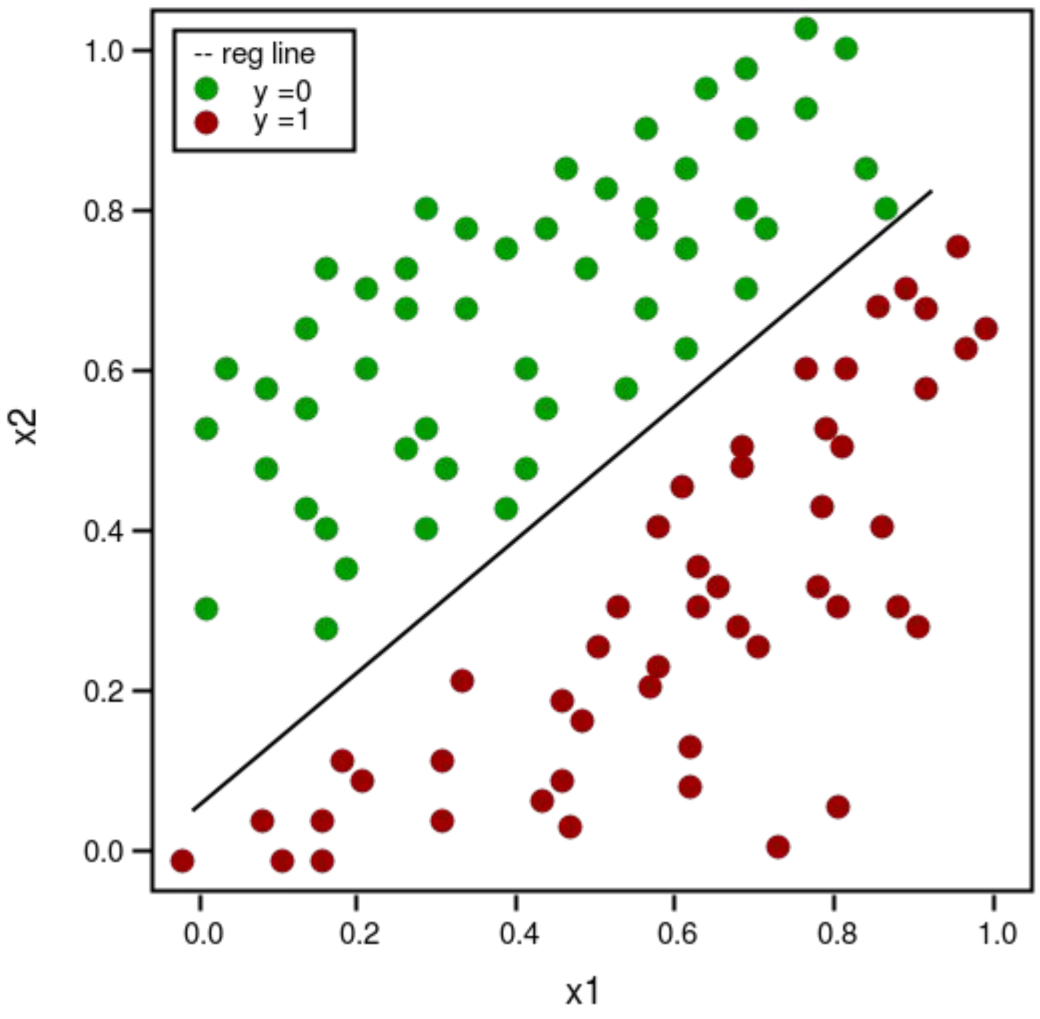
Logistic regression

$x_1, x_2 \in \mathbb{R}$ – inputs

$y \in \{0, 1\}$ – target

$w_0, w_1, w_2 \in \mathbb{R}$ – weights

$\hat{y} = f(w_0 + w_1 x_1 + w_2 x_2)$



Logistic regression

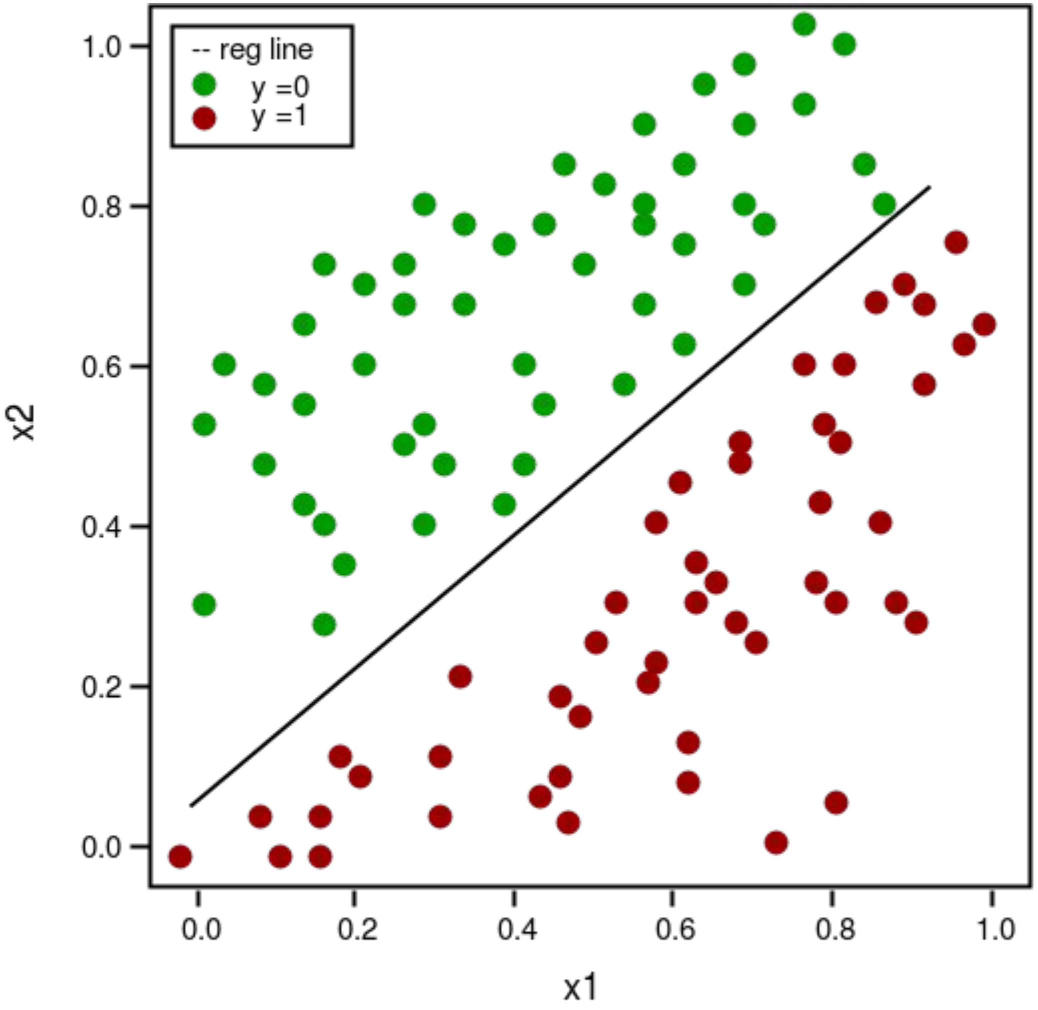
$x_1, x_2 \in \mathbb{R}$ – inputs

$y \in \{0, 1\}$ – target

$w_0, w_1, w_2 \in \mathbb{R}$ – weights

$$\hat{y} = f(w_0 + w_1 x_1 + w_2 x_2)$$

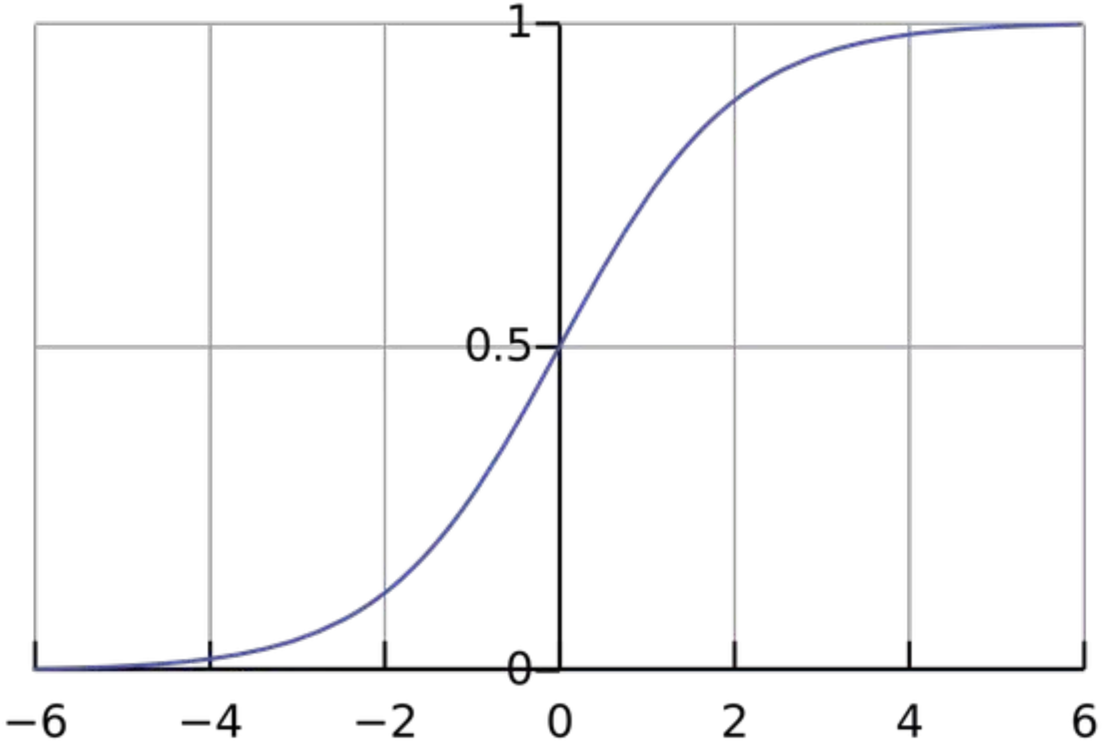
$$f(\cdot) : \mathbb{R} \longmapsto (0, 1)$$



Softmax function

$$\sigma(\cdot) : \mathbb{R} \longmapsto (0, 1)$$

$$\sigma(z) = \frac{1}{1 + exp(-z)}$$



Loss function: cross-entropy

$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

Loss function: cross-entropy

$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} ($$

Loss function: cross-entropy

$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \big(y_i \log \hat{y_i} +$$

Loss function: cross-entropy

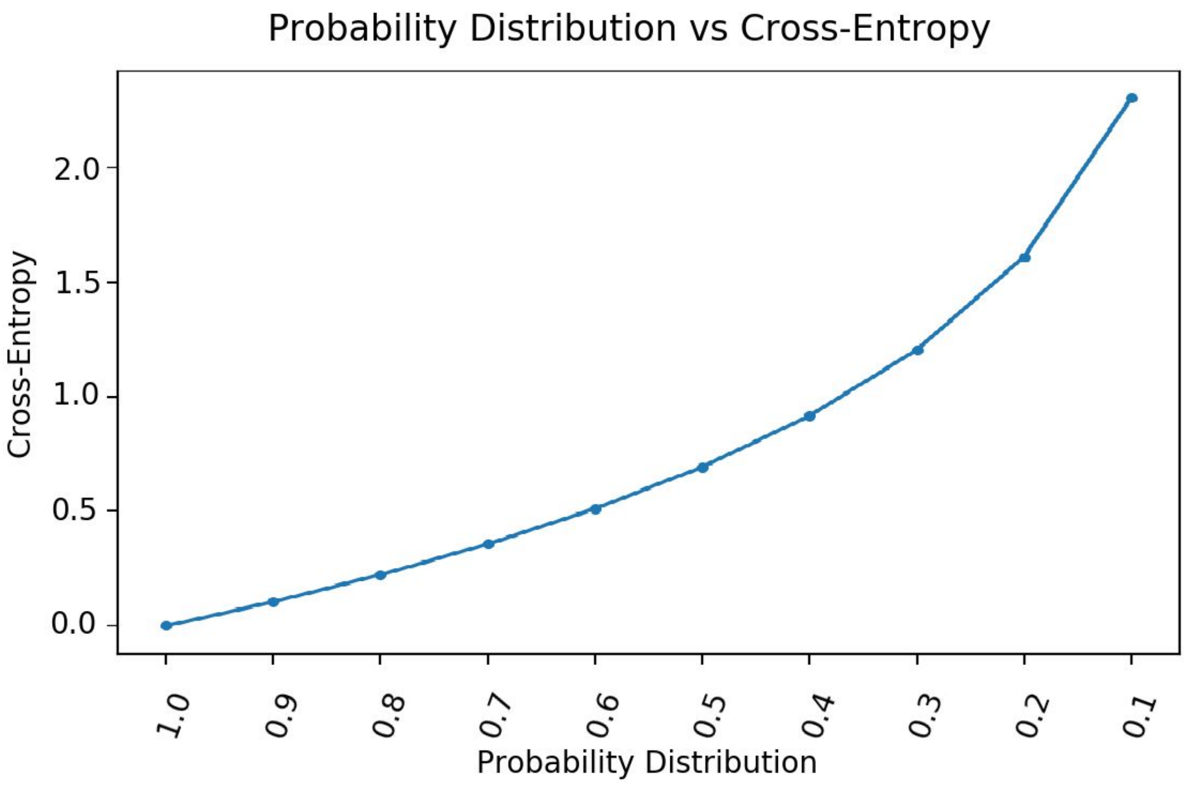
$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1 - y_i) \log (1 - \hat{y_i}) \right)$$

Loss function: cross-entropy

$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1 - y_i) \log (1 - \hat{y_i}) \right)$$

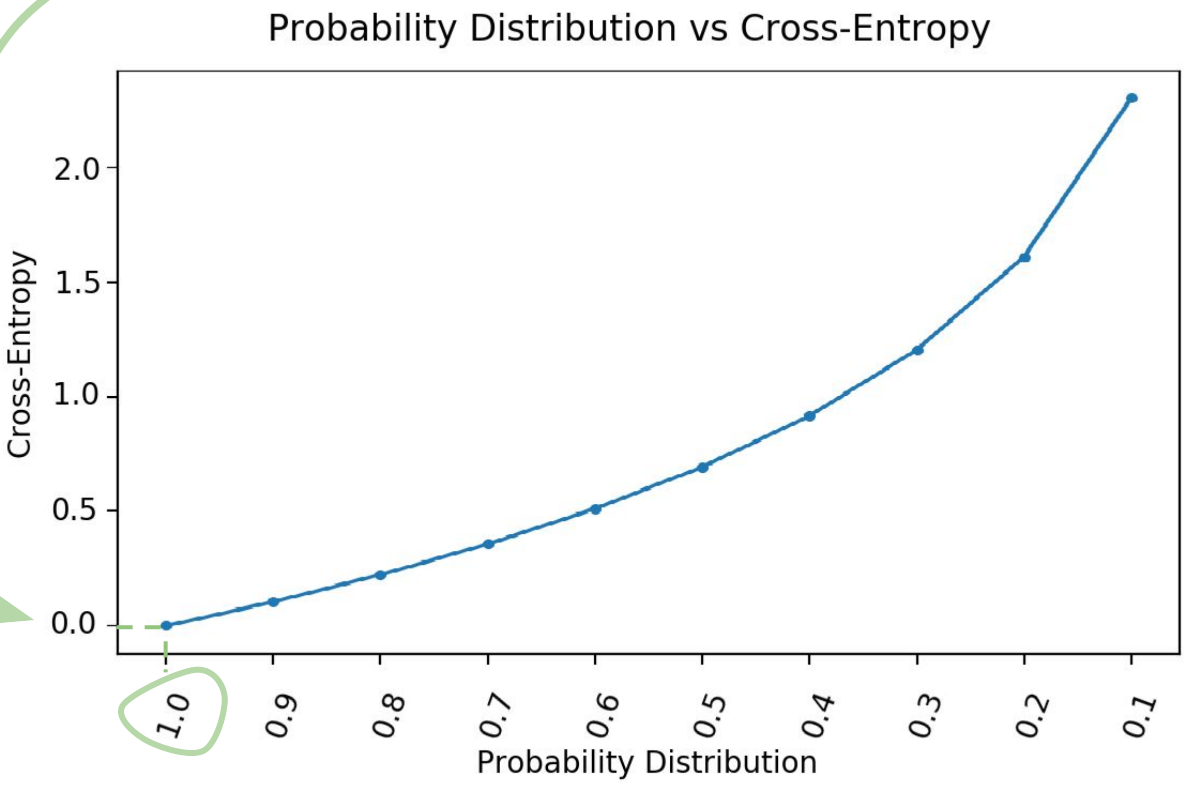


Loss function: cross-entropy

$y \in \{0, 1\}$ – target

$\hat{y} \in \mathbb{R}$ – prediction

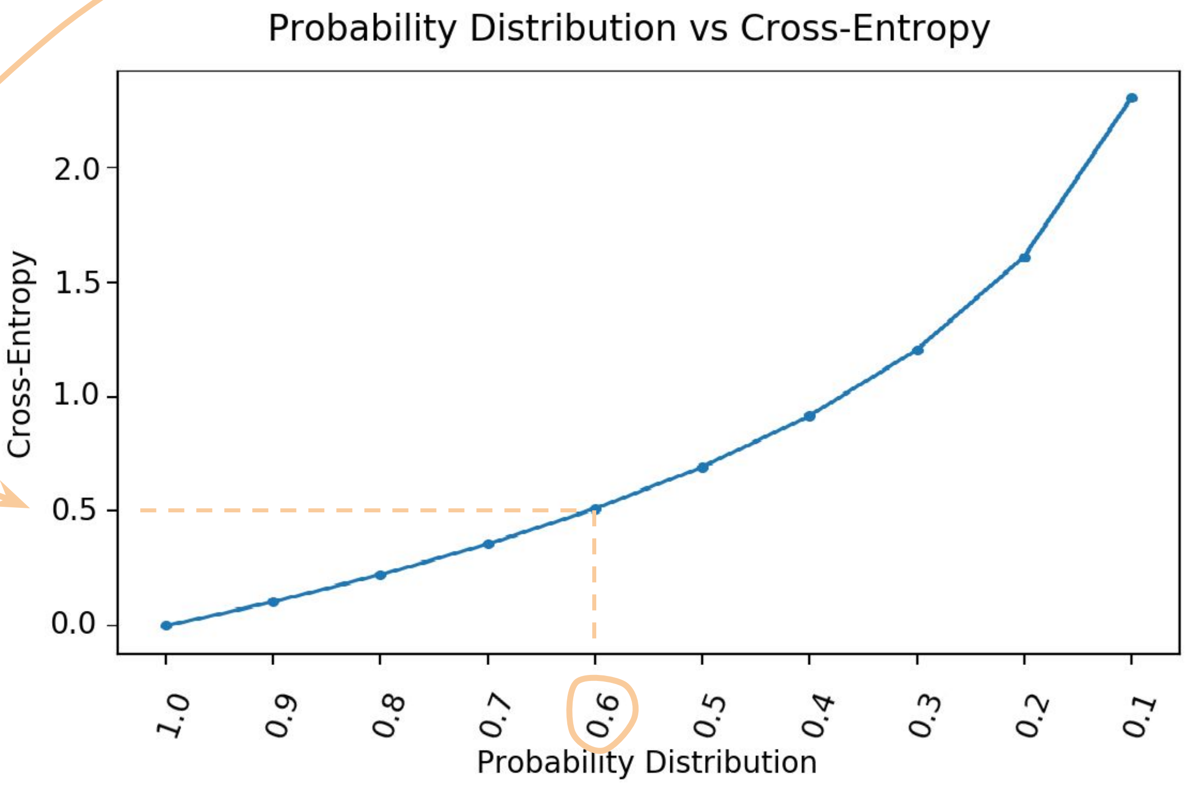
$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1 - y_i) \log (1 - \hat{y_i}) \right)$$



Loss function: cross-entropy

$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

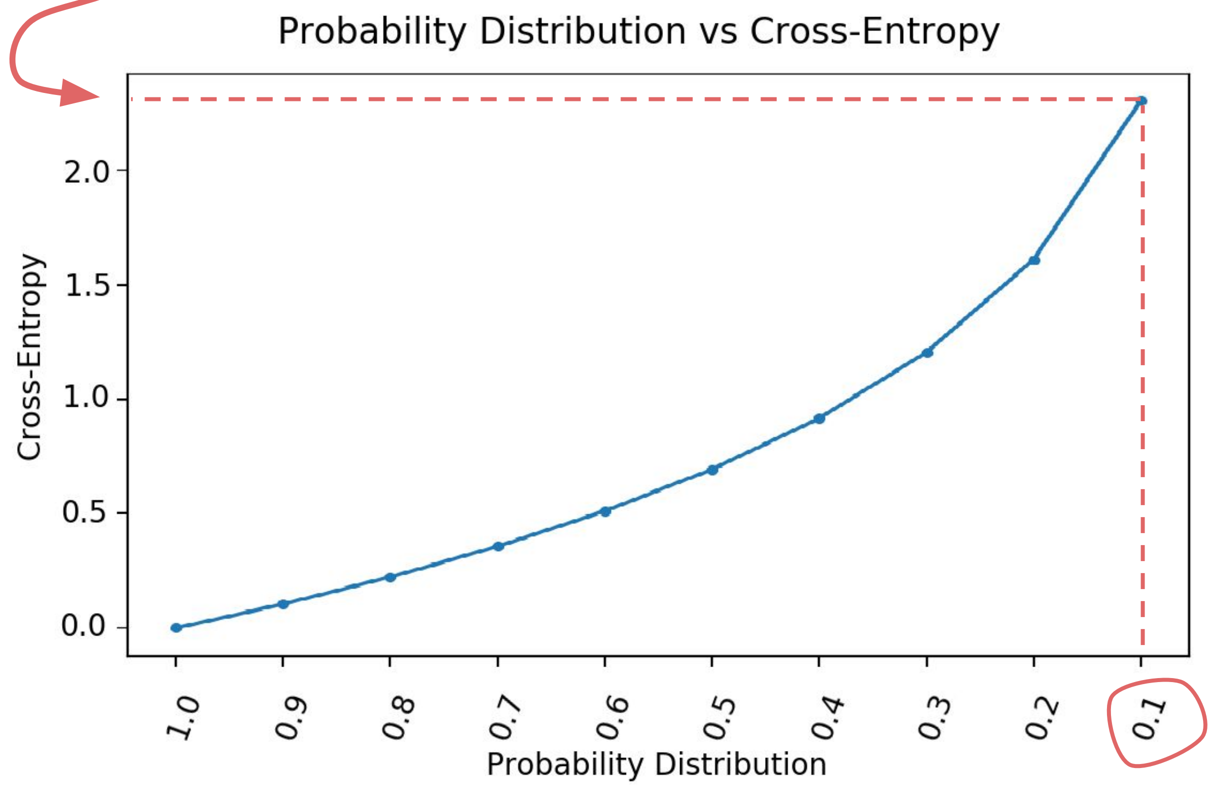
$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1 - y_i) \log (1 - \hat{y_i}) \right)$$



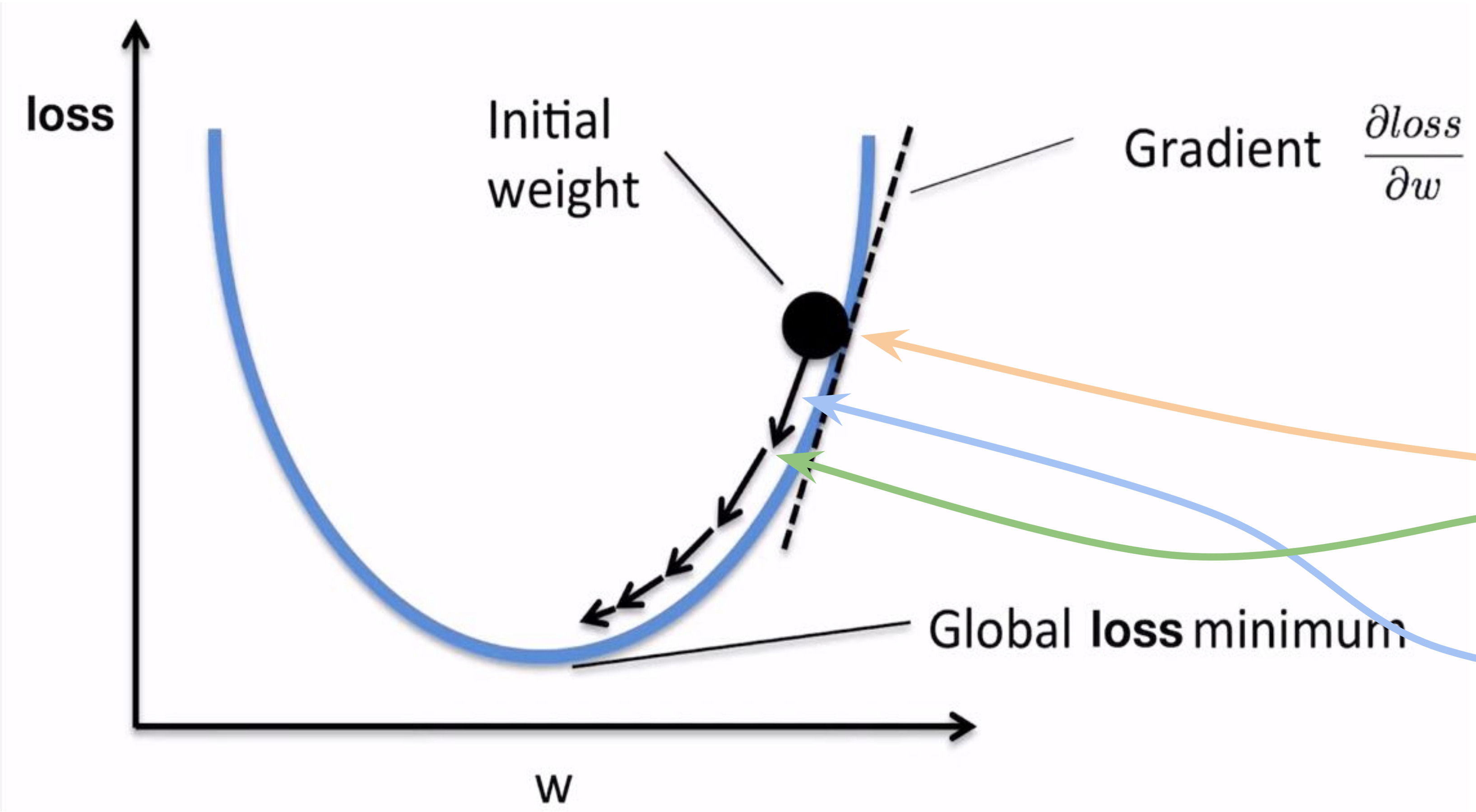
Loss function: cross-entropy

$y \in \{0, 1\}$ – target
$\hat{y} \in \mathbb{R}$ – prediction

$$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \left(y_i \log \hat{y_i} + (1 - y_i) \log (1 - \hat{y_i})\right)$$



Training process: gradient descent



$$\hat{y} = f(w_0 + w_1 x_1 + w_2 x_2)$$

$$L(y, \hat{y}) = \mathrm{CE}(y, \hat{y})$$

$$w_{new} = w_{old} - \alpha \frac{\partial L(w,x)}{\partial w}$$

$$\alpha \in \mathbb{R} - \text{learning rate}$$

Logistic regression overview

model equation: $$\hat{y} = \sigma\Big(\sum_{j=0}^{M} w_j g(x_j)\Big)$$

Logistic regression overview

model equation: $$\hat{y} = \sigma\Big(\sum_{j=0}^{M} w_j g(x_j)\Big)$$

loss function: $$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \big(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)\big)$$

Logistic regression overview

model equation: $$\hat{y} = \sigma\Big(\sum_{j=0}^{M} w_j g(x_j)\Big)$$

loss function: $$\mathrm{CE}(y, \hat{y}) = -\sum_{i=1}^{N} \big(y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)\big)$$

weights update: $$w_{new} = w_{old} - \frac{\partial L}{\partial w}$$

https://xkcd.com/1838/