Deep Learning
Summary of lecture 1

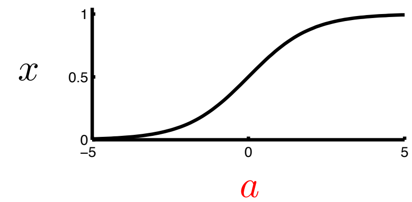
Dr. Richard E. Turner (ret26@cam.ac.uk)

Engineering Tripos Part IB
Paper 8: Information Engineering

Summary of lecture 1

single neuron

$$x(a) = \frac{1}{1+\exp(-a)}$$

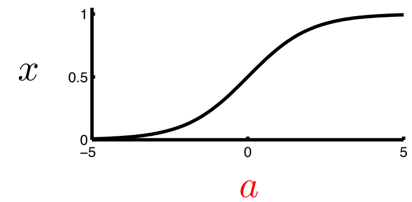


$$a = \sum_{d=0}^{D} w_d z_d$$

Summary of lecture 1

single neuron

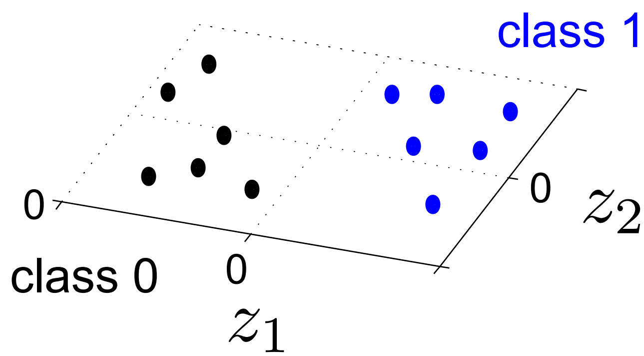
$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

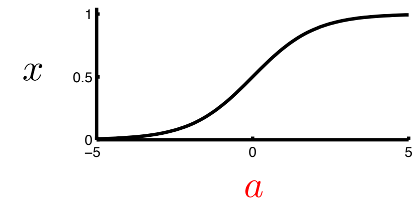
	inputs	class labels
training data	$\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$	$\{y^{(n)}\}_{n=1}^{N}$

Summary of lecture 1



single neuron

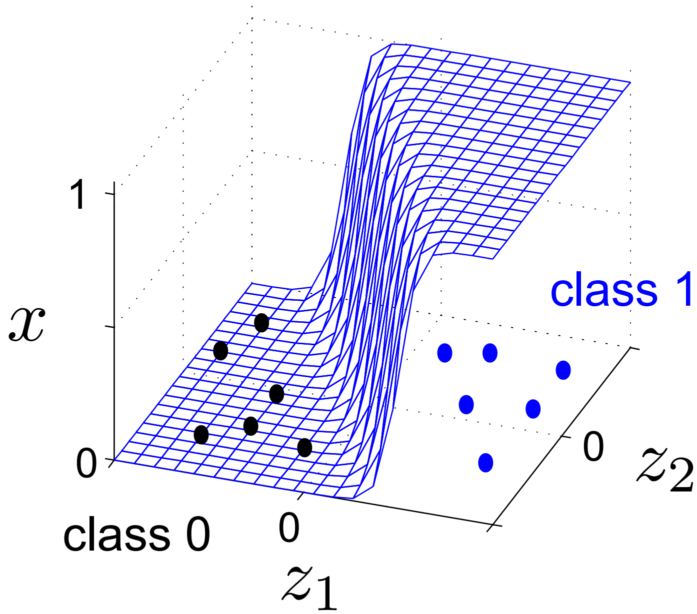
$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

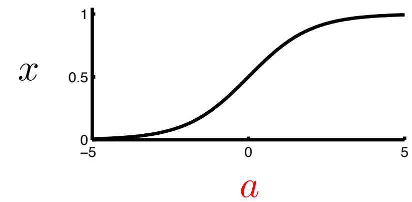
training data — inputs: $\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$ — class labels: $\{y^{(n)}\}_{n=1}^{N}$

Summary of lecture 1



single neuron

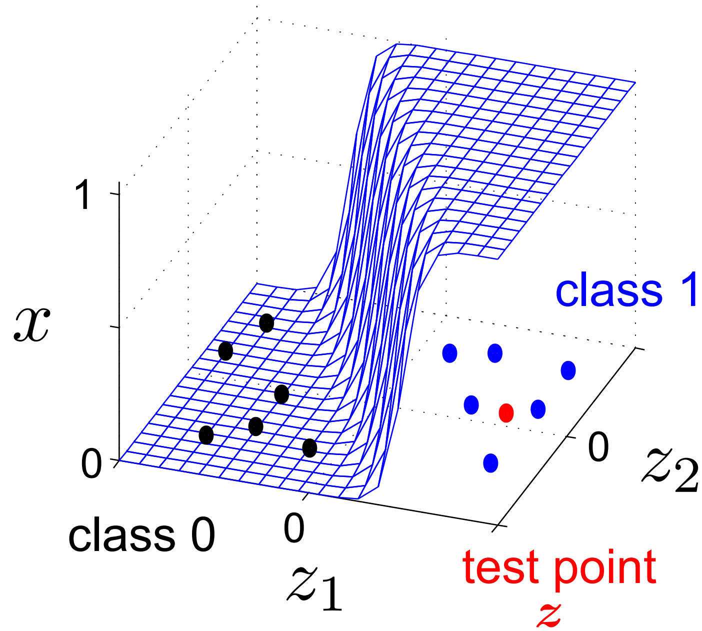
$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

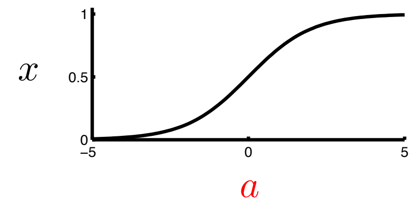
	inputs	class labels
training data	$\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$	$\{y^{(n)}\}_{n=1}^{N}$

Summary of lecture 1



single neuron

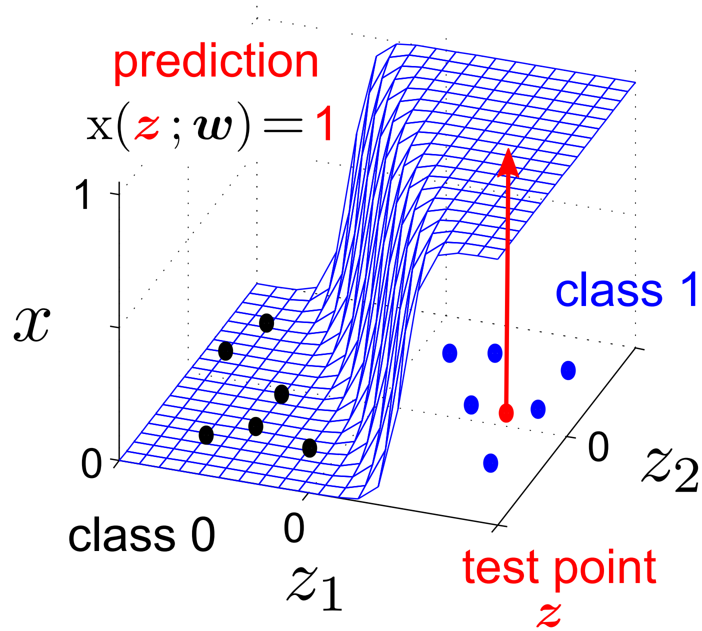
$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

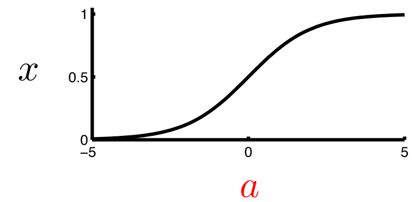
	inputs	class labels
training data	$\{\mathbf{z}^{(n)}\}_{n=1}^{N}$	$\{y^{(n)}\}_{n=1}^{N}$

Summary of lecture 1



single neuron

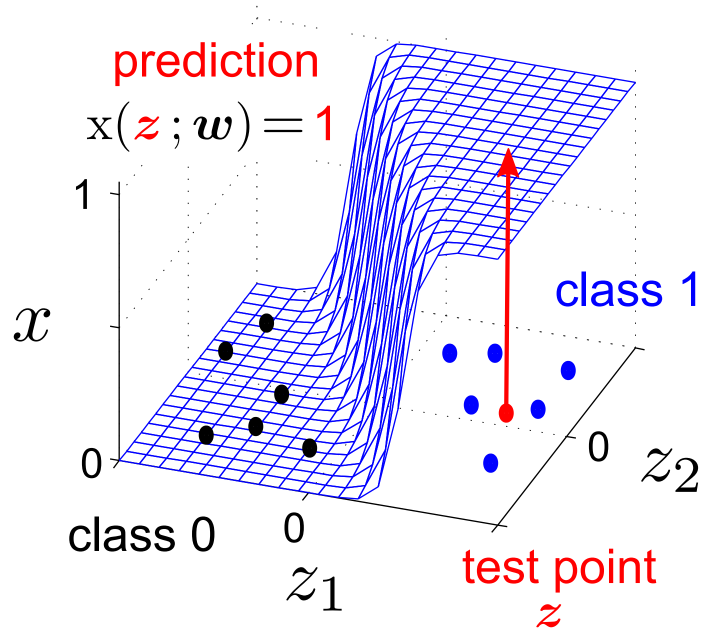
$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

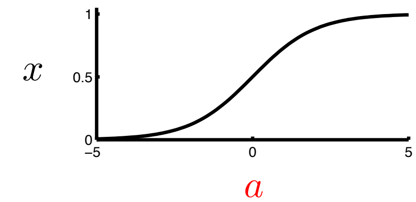
	inputs	class labels
training data	$\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$	$\{y^{(n)}\}_{n=1}^{N}$

Summary of lecture 1



single neuron

$$x(a) = \frac{1}{1+\exp(-a)}$$



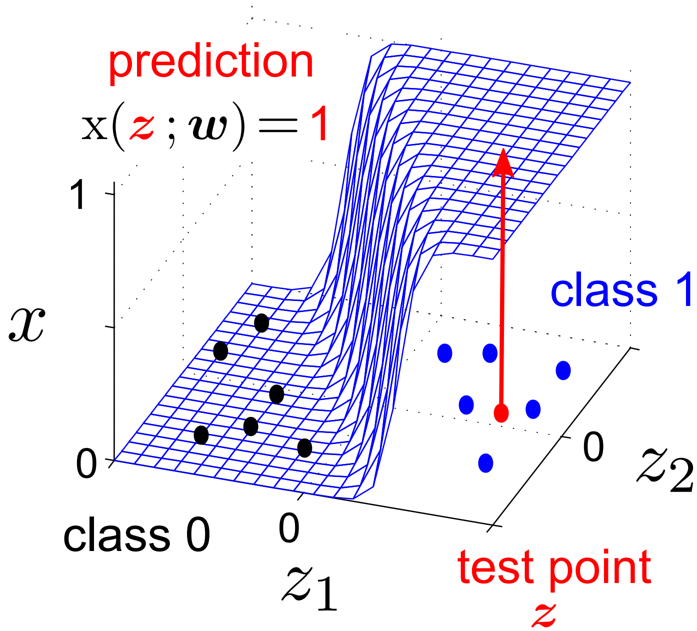
$$a = \sum_{d=0}^{D} w_d z_d$$

	inputs	class labels
training data	$\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$	$\{y^{(n)}\}_{n=1}^{N}$

objective function:

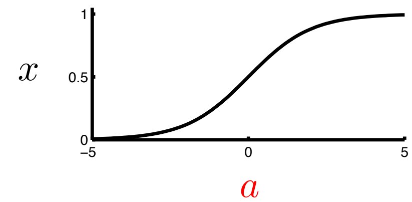
$$G(\boldsymbol{w}) = -\sum_n \left[y^{(n)} \log \mathrm{x}(\boldsymbol{z}^{(n)}; \boldsymbol{w}) + (1 - y^{(n)}) \log \left(1 - \mathrm{x}(\boldsymbol{z}^{(n)}; \boldsymbol{w})\right)\right] \geq 0$$

Summary of lecture 1



single neuron

$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

	inputs	class labels
training data	$\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$	$\{y^{(n)}\}_{n=1}^{N}$

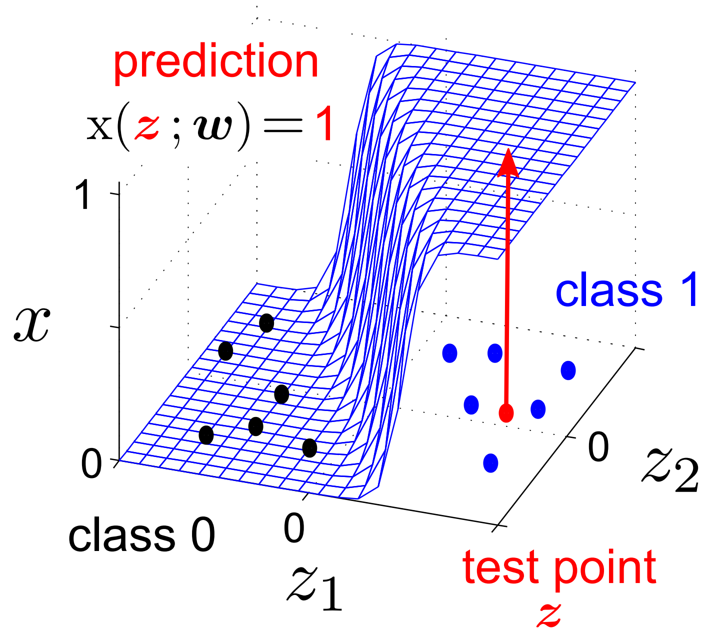
objective function:

$$G(\boldsymbol{w}) = -\sum_n \left[y^{(n)} \log \mathrm{x}(\boldsymbol{z}^{(n)};\boldsymbol{w}) + (1 - y^{(n)}) \log \left(1 - \mathrm{x}(\boldsymbol{z}^{(n)};\boldsymbol{w})\right) \right] \geq 0$$

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} G(\boldsymbol{w})$$

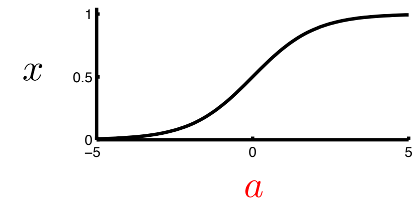
choose weights that minimise network's surprise about training data

Summary of lecture 1



single neuron

$$x(a) = \frac{1}{1+\exp(-a)}$$



$$a = \sum_{d=0}^{D} w_d z_d$$

training data — inputs $\{\boldsymbol{z}^{(n)}\}_{n=1}^{N}$ — class labels $\{y^{(n)}\}_{n=1}^{N}$

objective function:

$$G(\boldsymbol{w}) = -\sum_n \left[y^{(n)} \log \mathrm{x}(\boldsymbol{z}^{(n)}; \boldsymbol{w}) + (1 - y^{(n)}) \log \left(1 - \mathrm{x}(\boldsymbol{z}^{(n)}; \boldsymbol{w})\right)\right] \geq 0$$

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} G(\boldsymbol{w})$$ choose weights that minimise network's surprise about training data

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \eta \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{w}} G(\boldsymbol{w})$$ iteratively step down the objective (gradient points up hill)