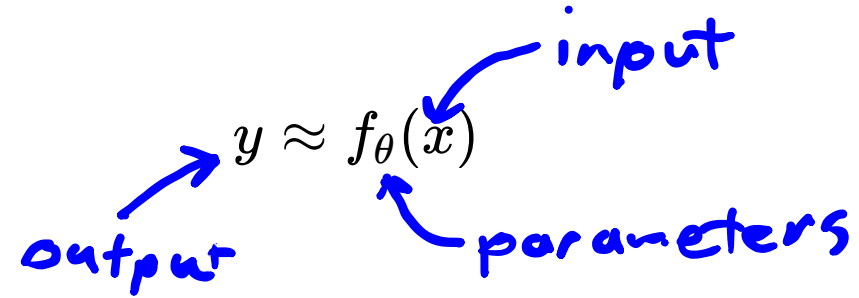# Map of RL Algorithms

# This Time

Challenges in Reinforcement Learning:

- Exploration vs Exploitation
- Credit Assignment
- Generalization
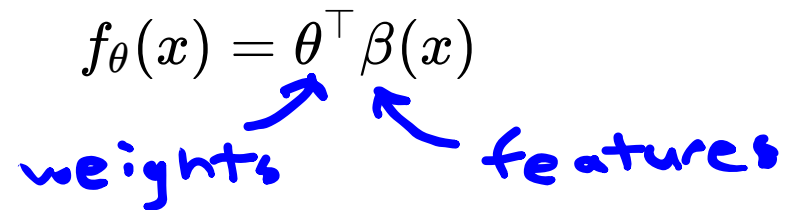
# Function Approximation

$$y \approx f_\theta(x)$$

input

output

parameters

Previously, Linear:

$$f_\theta(x) = \theta^\top \beta(x)$$

weights   features

e.g. $\beta_i(x) = \sin(i\,\pi\,x)$

# Neural Network

# Nonlinearities

| sigmoid | tanh | softplus |
|---|---|---|
| $1/(1 + \exp(-x))$ | $\tanh(x)$ | $\log(1 + \exp(x))$ |

| relu | leaky relu | swish |
|---|---|---|
| $\max(0, x)$ | $\max(\alpha x, x)$ | $x \, \text{sigmoid}(x)$ |
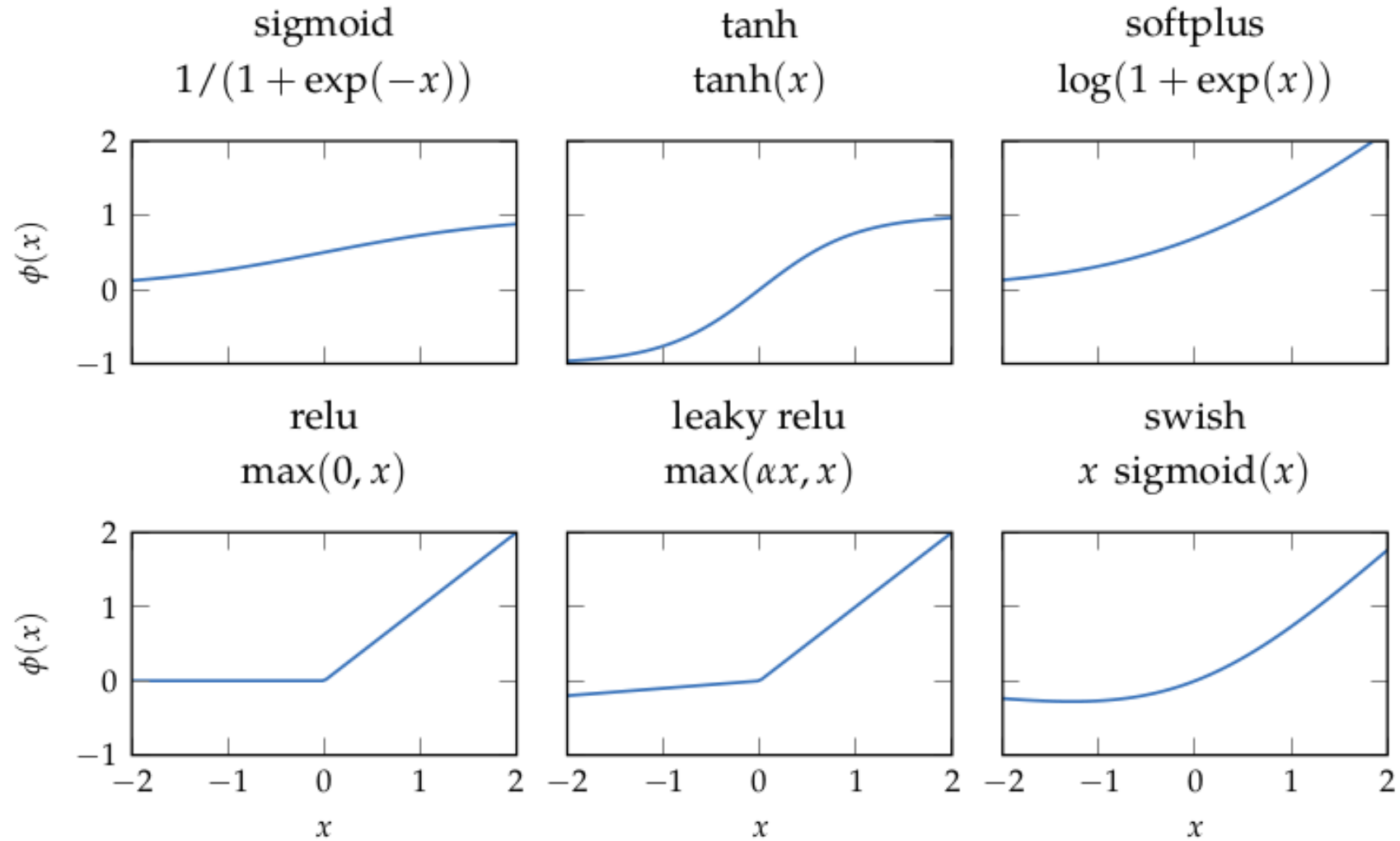
# Training
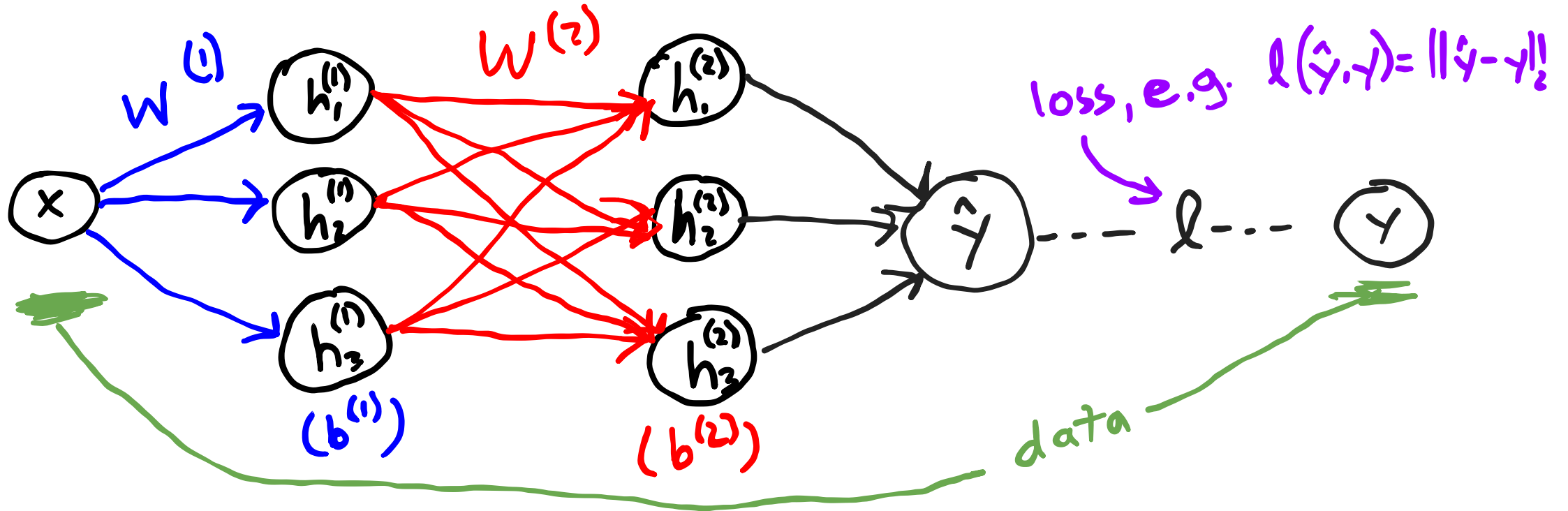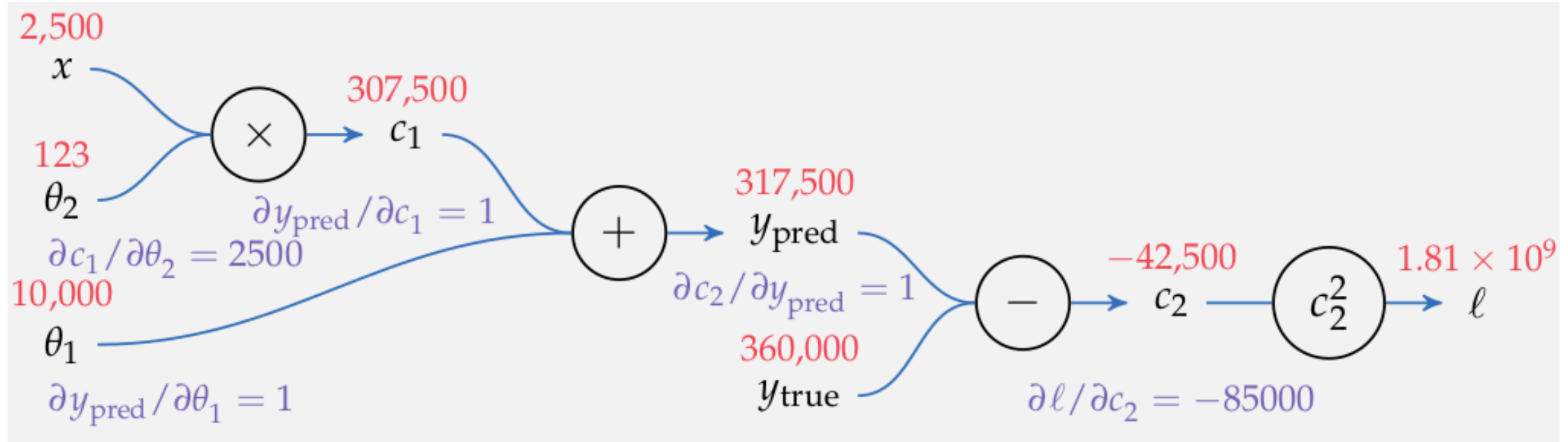


$$\theta^* = \arg\min_{\theta} \sum_{(x,y)\in\mathcal{D}} l(f_\theta(x), y)$$

Stochastic Gradient Descent: $\theta \leftarrow \theta - \alpha \nabla_\theta l(f_\theta(x), y)$

# Chain Rule

# Backprop



$$\frac{\partial \ell}{\partial \theta_1} = \frac{\partial \ell}{\partial c_2} \frac{\partial c_2}{\partial y_{\text{pred}}} \frac{\partial y_{\text{pred}}}{\partial \theta_1} = -85{,}000 \cdot 1 \cdot 1 = -85{,}000$$

$$\frac{\partial \ell}{\partial \theta_2} = \frac{\partial \ell}{\partial c_2} \frac{\partial c_2}{\partial y_{\text{pred}}} \frac{\partial y_{\text{pred}}}{\partial c_1} \frac{\partial c_1}{\partial \theta_2} = -85{,}000 \cdot 1 \cdot 1 \cdot 2500 = -2.125 \times 10^8$$

# Adaptive Step Size: RMSProp
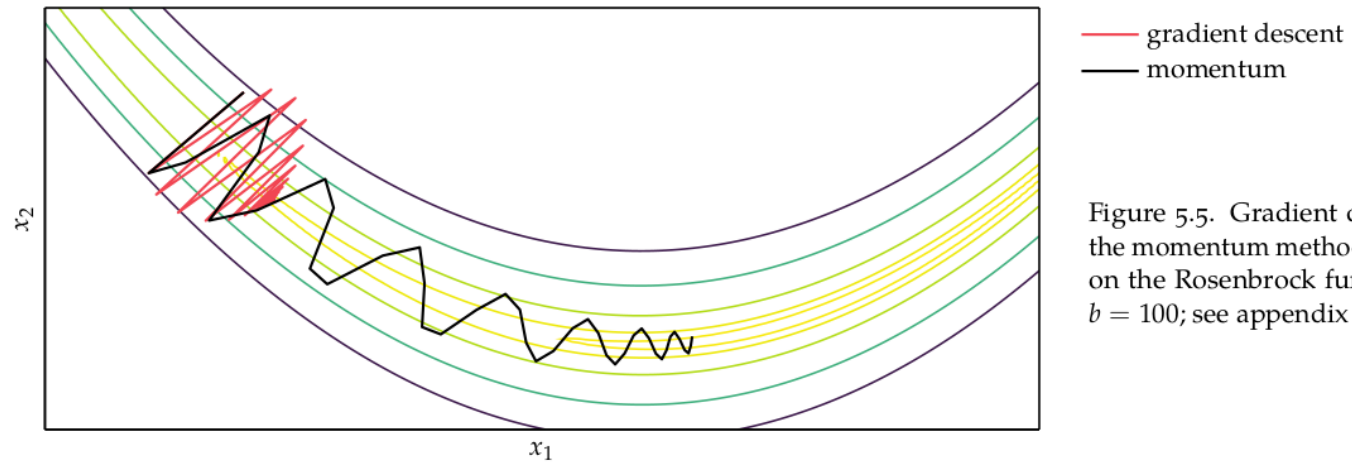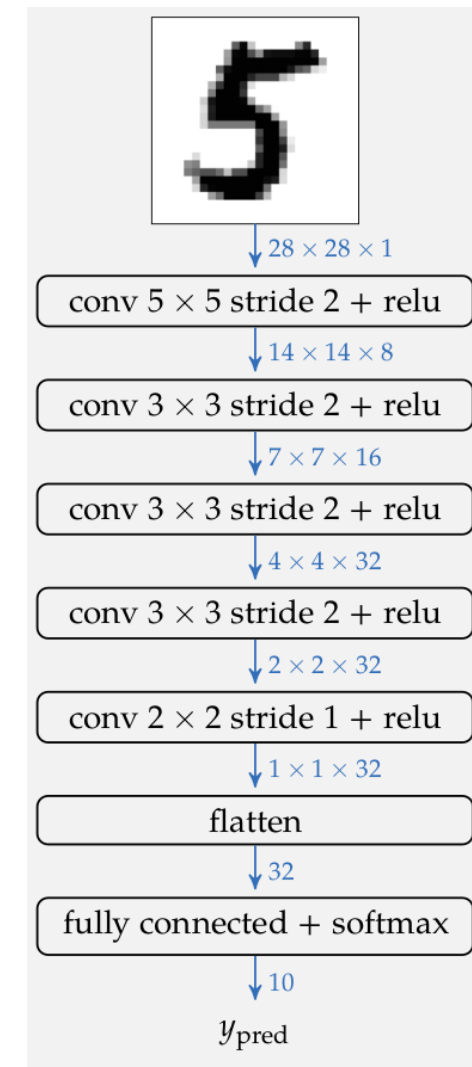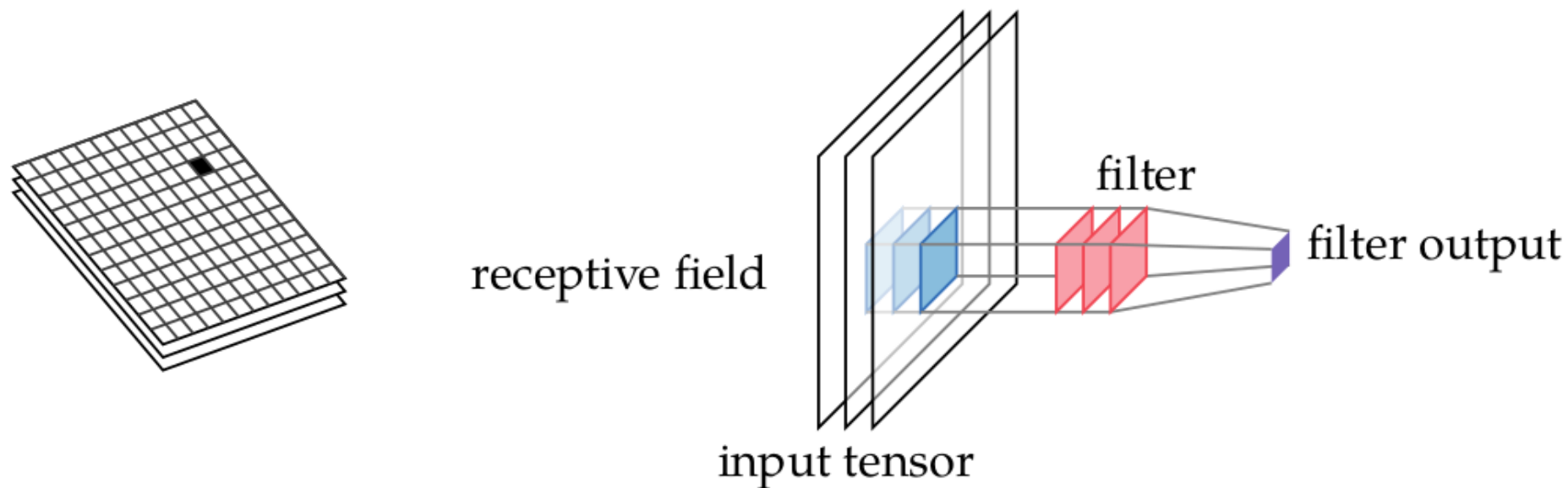
# Adaptive Step Size: ADAM



Figure 5.5. Gradient descent and the momentum method compared on the Rosenbrock function with $b = 100$; see appendix B.6.

# On Your Radar: ConvNets



receptive field

input tensor

filter

filter output

$28 \times 28 \times 1$

conv $5 \times 5$ stride 2 + relu

$14 \times 14 \times 8$

conv $3 \times 3$ stride 2 + relu

$7 \times 7 \times 16$

conv $3 \times 3$ stride 2 + relu

$4 \times 4 \times 32$

conv $3 \times 3$ stride 2 + relu

$2 \times 2 \times 32$

conv $2 \times 2$ stride 1 + relu

$1 \times 1 \times 32$

flatten

$32$

fully connected + softmax

$10$

$y_{\text{pred}}$

# On Your Radar: Regularization

$$\arg\min_{\boldsymbol{\theta}} \sum_{(x,y)\in\mathbf{D}} \ell(f_{\boldsymbol{\theta}}(x),y) - \beta\|\boldsymbol{\theta}\|^2$$

e.g. Batch norm, dropout