# Overview

# Stochastic Gradient Descent (SGD)

## Algorithm

Update step:

$$\theta_{t+1} = \theta_t - \eta \cdot \nabla_\theta J(\theta_t) \tag{1}$$

# SGD with Momentum

## Algorithm

Update step:

$$v_{t,i} = \gamma \cdot v_{t-1,i} + \nabla_\theta J(\theta_{t,i}) \tag{2}$$

$$\theta_{t+1} = \theta_t - \eta \cdot v_{t,i} \tag{3}$$

# Optimizers in PyTorch

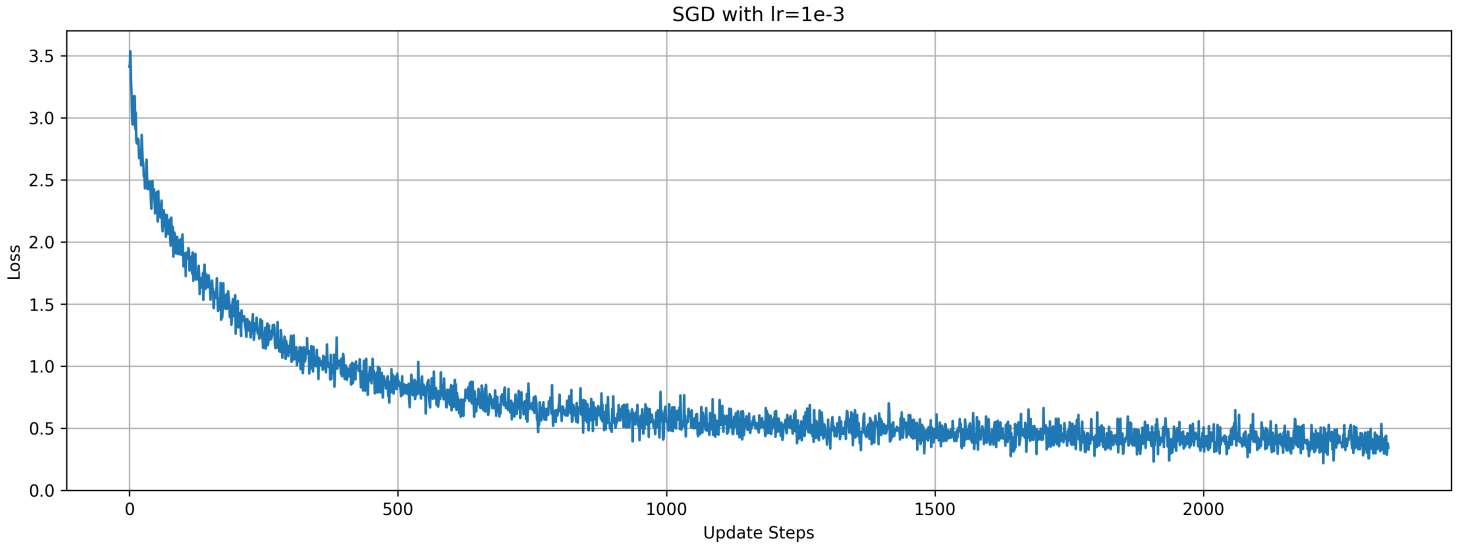
Vanilla training loop in PyTorch:

```
for input, target in dataset:
  optimizer.zero_grad()
  output = model(input)
  loss = loss_fn(output, target)
  loss.backward()
  optimizer.step()
```

How to use an optimizer:

```
optimizer = optim.SGD(model.parameters(), lr=0.01, momentum=0.9)
#optimizer = optim.Adam(model.parameters(), lr=0.0001)
```

# First images in beamer



# Citation

An example of the `\cite` command to cite within the presentation:

This statement requires citation. [Tob63]

# References I


James Tobin, *Commercial banks as creators of 'money'*, no. 159.