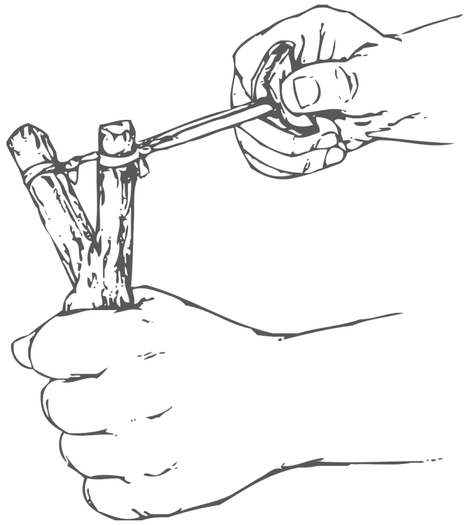
Neural·Pragmatic
Natural
Language
Generation

N·P
NLG

Learning goals



1. Understand the basics of Gradient Descent.
2. Get familiar with variations of GD.
3. Learn how to us GD to optimize an RSA model.

Stochastic Gradient Descent

input : γ (lr), θ_0 (params), $f(\theta)$ (objective), λ (weight decay),
μ (momentum), τ (dampening), *nesterov*, *maximize*

for $t = 1$ **to** ... **do**

$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$

if $\lambda \neq 0$

$g_t \leftarrow g_t + \lambda\theta_{t-1}$

if $\mu \neq 0$

if $t > 1$

$\mathbf{b}_t \leftarrow \mu\mathbf{b}_{t-1} + (1-\tau)g_t$

else

$\mathbf{b}_t \leftarrow g_t$

if *nesterov*

$g_t \leftarrow g_t + \mu\mathbf{b}_t$

else

$g_t \leftarrow \mathbf{b}_t$

if *maximize*

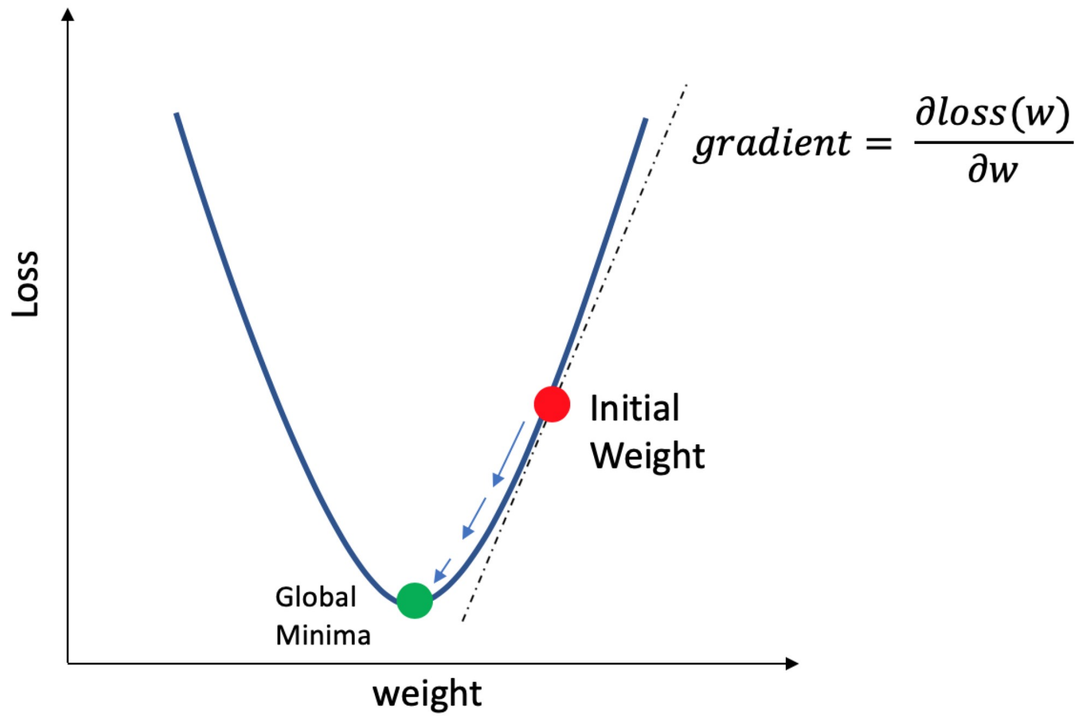
$\theta_t \leftarrow \theta_{t-1} + \gamma g_t$

else

$\theta_t \leftarrow \theta_{t-1} - \gamma g_t$

return θ_t

from PyTorch Documentation of SGD



from paper

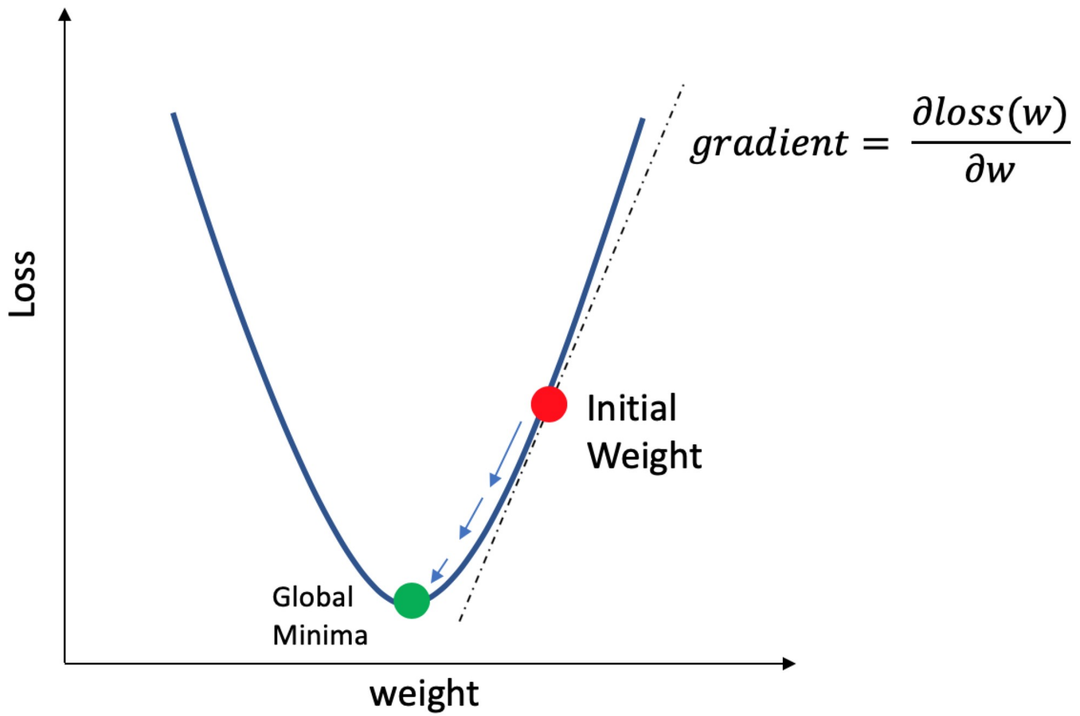
Stochastic Gradient Descent

input : γ (lr), θ_0 (params), $f(\theta)$ (objective), λ (weight decay),
μ (momentum), τ (dampening), *nesterov*, *maximize*

for $t = 1$ **to** $\ldots$ **do**
  $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$
  if $\lambda \neq 0$
    $g_t \leftarrow g_t + \lambda\theta_{t-1}$
  if $\mu \neq 0$
    if $t > 1$
      $\mathbf{b}_t \leftarrow \mu\mathbf{b}_{t-1} + (1-\tau)g_t$
    else
      $\mathbf{b}_t \leftarrow g_t$
    if *nesterov*
      $g_t \leftarrow g_t + \mu\mathbf{b}_t$
    else
      $g_t \leftarrow \mathbf{b}_t$
  if *maximize*
    $\theta_t \leftarrow \theta_{t-1} + \gamma g_t$
  else
    $\theta_t \leftarrow \theta_{t-1} - \gamma g_t$

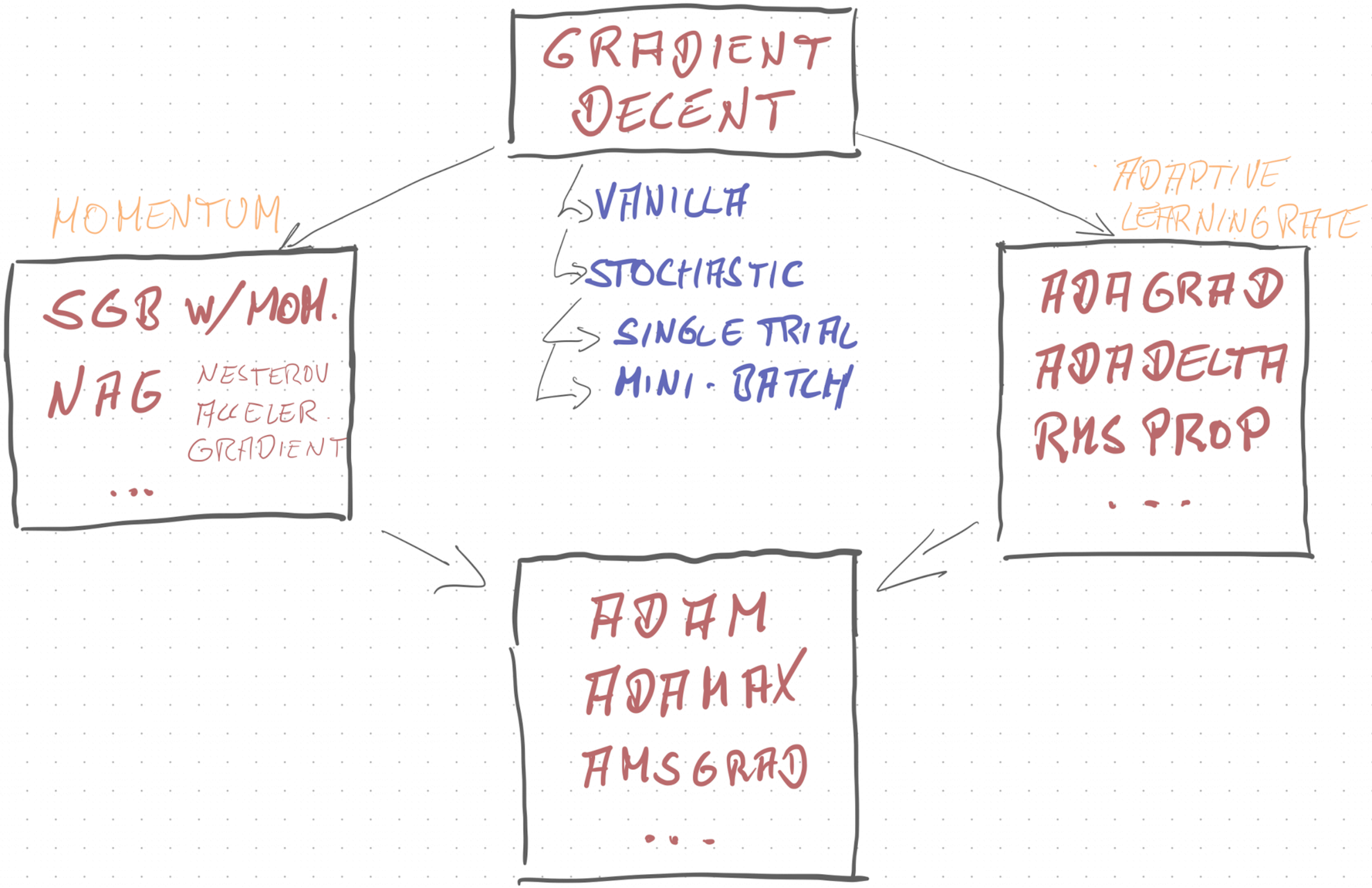
return θ_t

from PyTorch Documentation of SGD



from paper

Common optimization algorithms



Rational Speech Act (RSA) model

LITERAL INTERPRETATION

$$P_{lit}(s \mid u) \propto \mathscr{L}(u, s)$$

			
“square”	.5	0	.5
“circle”	0	1	0
“green”	0	0	1
“blue”	.5	.5	0

PRAGMATIC SPEAKER

$$P_S(u \mid s) = \mathrm{SM}\left(\log P_{lit}(s \mid u) - \mathrm{C}(u)\right)$$

	“square”	**“circle”**	**“green”**	**“blue”**
	.5	0	0	.5
	0	.89	0	.11
	.11	0	.89	0

PRAGMATIC INTERPRETATION

$$P_L(s \mid u) \propto P_{sal}(s)\ P_S(u \mid s)$$

			
“square”	.82	0	.18
“circle”	0	1	0
“green”	0	0	1
“blue”	.82	.18	0

Human experiments with reference games

1. speaker production condition

   choose an utterance for a given referent

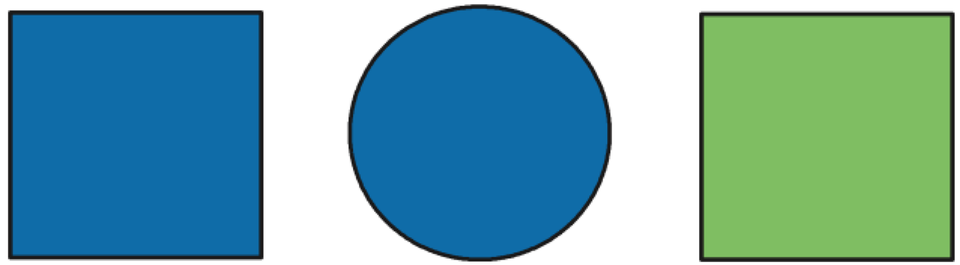
2. listener interpretation condition

   choose a referent for a given utterance

3. salience prior elicitation condition

   choose a referent for a given utterance you do not know the meaning of

$$P_L(s \mid u) \propto P_{sal}(s)\ P_S(u \mid s)$$

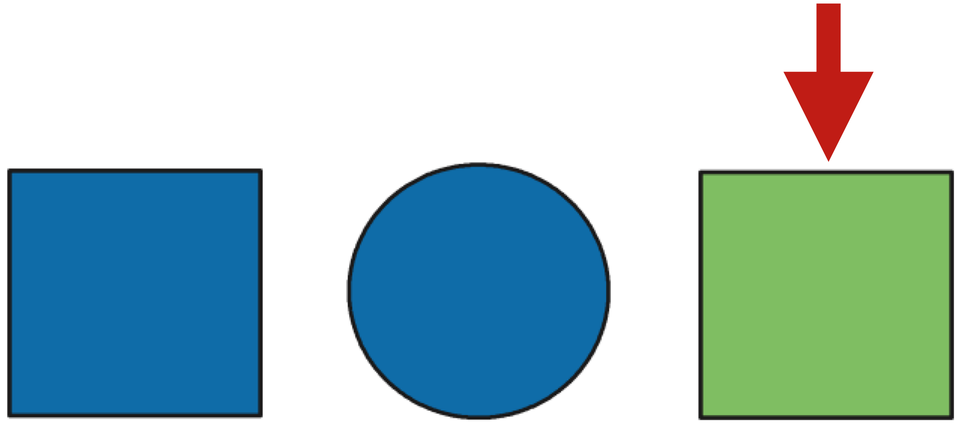


circle square

green blue

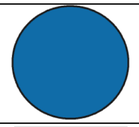
Speaker production condition

choose an utterance for a given referent



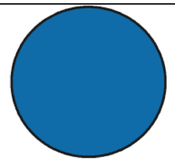
circle square

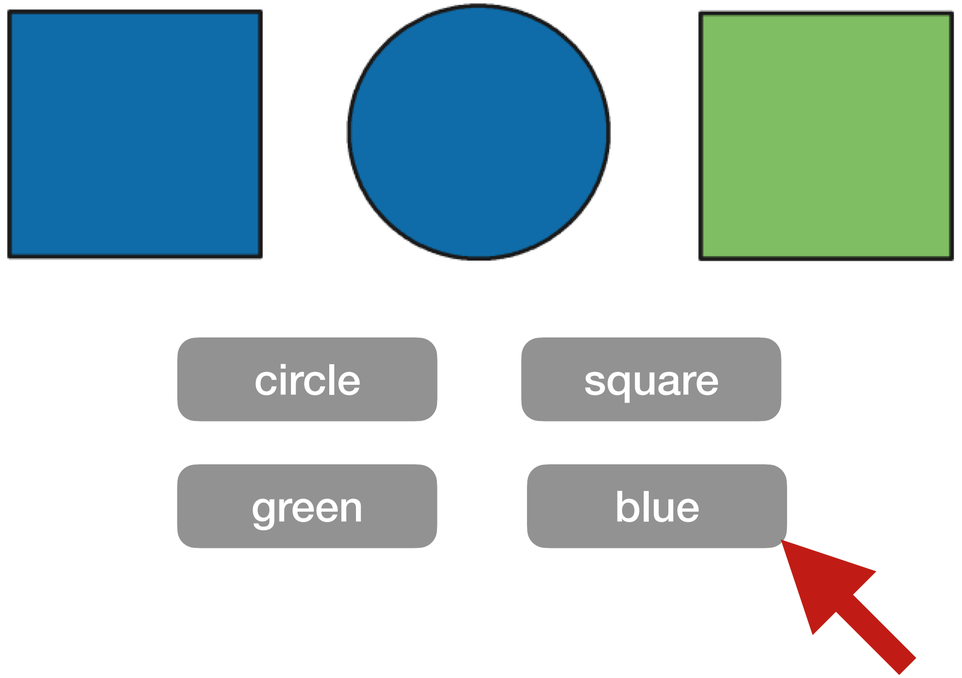
green blue

	blue	circle	green	square
	9	135	0	0
	0	0	119	25
	63	0	0	81

Listener interpretation condition

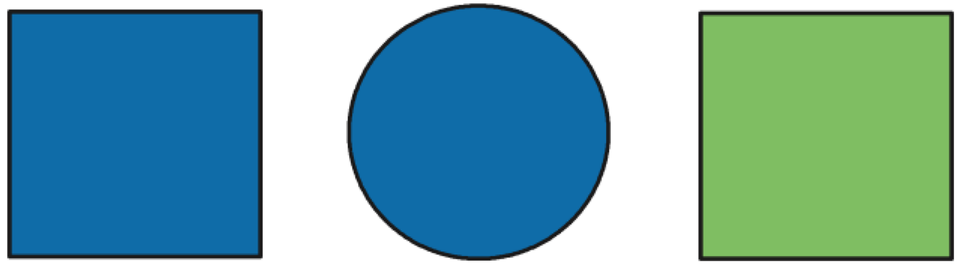
choose a referent for a given utterance

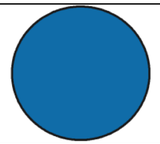
			
blue	66	0	115
square	0	117	62



Salience prior elicitation condition

choose a referent for a given utterance you do not know the meaning of



			
blue	71	139	30

$$P_L(s \mid u) \propto P_{sal}(s)\ P_S(u \mid s)$$