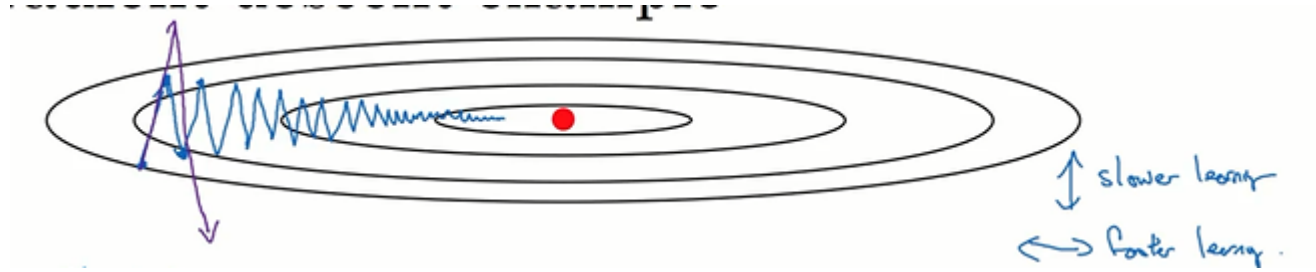
Gradient decsent momentum



- in horizontal direction, we want the learning go fast
- in vertical direction, we want the learning go slow, not oscillating too much, thus slow down the whole learning process

---

Momentum

$v_{dw} = 0,\ v_{db} = 0$

$On\ iteration\ t:$

$compute\ dW,\ db\ on\ mini-batch$

$v_{dW} = \beta v_{dW} + (1-\beta)dW$

$v_{db} = \beta v_{db} + (1-\beta)db$

$W := W - \alpha v_{dW}$

$b := b - \alpha v_{db}$

$Hyperparameters: \alpha,\ \beta$

$\beta = 0.9$

- *Analogy:*
  - assume a ball-shaped cost function where we roll down to the minimum
  - dW, db are acceleration
  - v_{dW}, v_{db} are velocity