**Lec 6 polls**

**Slide 30**

Minimizing the (differentiable) loss function will also minimize classification error, true or false

* True
* False

**Slide 103**

Mark all true statements

* Step sizes that are greater than twice the inverse of the second derivative can cause gradient descent to diverge
* This is always a bad thing
* Gradient descent will not converge without decaying learning rates

**Slide 116**

The derivative of the loss w.r.t a parameter w, computed at the current estimate is positive. After taking a step (updating the parameter by a increment dw) the sign of the derivative becomes negative. Mark all true statements

* Rprop will revert to the earlier estimate and take a smaller step
* Rprop will change direction and begin taking steps in the opposite direction

**Slide 140**

On a flat surface of constant slope momentum methods will converge faster than vanilla gradient descent, true or false

* True
* False – momentum only changes step size