CS182 HW1.

1.(a). $w_{t+1} - w^* = (1-2\eta\sigma^2)(w_t - w^*)$

$\therefore \quad w^* = \frac{y}{\sigma}$

$\therefore \quad w_t = w^* - w^*(1-2\eta\sigma^2)^t$

$\because$ stable

$\therefore \quad |1-2\eta\sigma^2| < 1$

$\therefore \quad -1 < 1-2\eta\sigma^2 < 1$

$\therefore \quad 0 < \eta < \frac{1}{\sigma^2}$, converge to w^*.

(b). $w_t = (1-(1-2\eta\sigma^2)^t)\, w^*$

$w_t - w^* = (1-2\eta\sigma^2)^t \cdot w^*$

$\therefore \quad w^*(1-2\eta\sigma^2)^t \le \varepsilon \cdot w^*$

$\therefore \quad (1-2\eta\sigma^2)^t \le \varepsilon$

$\therefore \quad t \geq \frac{\log \varepsilon}{\log |1-2\eta\sigma^2|}$

$\therefore \quad t_{min} = \left\lceil \frac{\log \varepsilon}{\log |1-2\eta\sigma^2|} \right\rceil$

(c). $\begin{cases} \sigma_\ell w[1] = y[1] \\ \sigma_s w[2] = y[2] \end{cases}$

$\therefore \quad |1-2\eta\sigma_\ell^2| < 1, \quad |1-2\eta\sigma_s^2| < 1$

$\therefore \quad \eta < \frac{1}{\sigma_\ell^2}, \quad \eta < \frac{1}{\sigma_s^2}$

$\because \quad \sigma_\ell >> \sigma_s$

$\therefore \quad \eta < \frac{1}{\sigma_\ell^2}$

(d). σ_ℓ faster.

4/(a). A. $\beta_1 \theta_{t-1} + (1-\beta_1) g_t$

B. $\beta_2 v_{t-1} + (1-\beta_2) g_t^2$

(b). $\nabla f_t^{reg}(\theta) = \nabla f_t(\theta) + \lambda\theta$

$$\theta_t = \theta_{t-1} - \eta \nabla f_{t-1}^{reg}(\theta_{t-1}) = \theta_{t-1} - \eta(\nabla f_{t-1}(\theta_{t-1}) + \lambda\theta_{t-1})$$

$$= (1-\eta\lambda)\theta_{t-1} - \eta \nabla f_{t-1}(\theta_{t-1})$$

$$\theta_t = (1-\gamma)\theta_{t-1} - \eta \nabla f_{t-1}(\theta_{t-1})$$

identical if $1-\gamma = 1-\eta\lambda$.

i.e. $\gamma = \eta\lambda$

$\therefore$ SGD with weight decay using $\gamma = \eta\lambda$ on $f_t(\theta)$ is equal to vanilla SGD on L2-regularized loss $f_t^{reg}(\theta)$.