[redacted]

(B) "Online gradient descent on the squared loss"

$$\ell(y,\hat{y}) = (y-\hat{y})^2$$

$$\frac{\partial}{\partial \hat{y}} \ell(y,\hat{y}) = 2(\hat{y}-y)$$

$$\nabla_w \ell(y, \langle w,x\rangle) = 2(\langle w,x\rangle - y)x$$

$(w_0 = 0)$

$$w_t = w_{t-1} - \eta_t \nabla_w \ell(y_t, \langle w, x_t\rangle)$$

$$= w_{t-1} - 2\eta_t(\langle w, x_t\rangle - y_t)\, x_t$$

~~Telescope~~

Realizable case: (w/out noise)

$$y_t = \langle w^*, x_t\rangle \quad \forall t$$

Then $\quad w_t = w_{t-1} - 2\eta_t \langle w - w^*, x\rangle x$

$$\|w^* - w_t\|^2 = \|w^* - w_{t-1} + 2\eta_t\langle w_{t-1} - w^*, x_t\rangle x_t\|^2$$

$(*)$
$$= \|w^* - w_{t-1}\|^2 - 4\eta_t\langle w^* - w_{t-1}, x_t\rangle^2$$

$$+ 4\eta_t^2\langle w^* - w_{t-1}, x_t\rangle^2 \|x_t\|^2$$

This is really good if we set $\eta_t = \eta \quad \forall t$

(Rmk: $*$ is saying "everytime we make a big mistake, we learn about w^*")

Because

$$\text{Regret} = \sum_t (y_t - \langle w_{t-1}, x_t\rangle)^2$$

$$= \sum_t \langle w^* - w_{t-1}, x_t\rangle^2$$

looks similar to rhs of $(*)$!

Telescope $(*)$

regret!

$$0 \le \|w^* - w_T\|^2 = \|w^* - w_0\|^2$$

$$- 4\eta\sum_t \langle w^* - w_{t-1}, x_t\rangle^2$$

$$+ 4\eta^2 \sum_t \langle w^* - w_{t-1}, x_t\rangle^2 \|x_t\|^2$$

$$\text{Regret} = \sum_t \langle w^* - w_{t-1}, x_t\rangle^2$$

$$\le \frac{1}{4\eta}\|w^* - w_0\|^2 + \eta\left(\sum_t \langle w^* - w_{t-1}, x_t\rangle^2 \|x_t\|^2\right)$$

Suppose $\|x_t\|^2 \le R^2$

$$\le \frac{1}{4\eta}\|w^* - w_0\|^2 + \eta R^2\left(\sum_t \langle w^* - w_{t-1}, x_t\rangle^2\right)$$

Rearrange ~~Require~~ $\eta \le \frac{1}{2R^2}$

$$\longrightarrow \text{Regret} \le \frac{1}{2\eta}\|w^* - w_0\|^2 = R^2\|w^* - w_0\|^2$$

With noise:

Independent mean-zero noise
(or M.D. Sequence)

~~Forsimplicity~~

$$y_t = \langle w^*, x_t \rangle + \xi_t$$

$$w_t = w_{t-1} - \eta \nabla_w \ell(y_t, \langle w_{t-1}, x_t \rangle)$$
$$= w_{t-1} - 2\eta (\langle w_{t-1} - w^*, x_t \rangle - \xi_t) x_t$$

so

$$\|w_t - w^*\|^2$$
$$= \|w_{t-1} - w^* - 2\eta(\langle w_{t-1} - w^*, x_t \rangle - \xi_t) x_t\|^2$$
$$= \|w_{t-1} - w^*\|^2 - 4\eta(\langle w_{t-1} - w^*, x_t\rangle - \xi_t)\langle w_{t-1} - w^*, x_t\rangle$$
$$+ 4\eta^2 (\langle w_{t-1} - w^*, x_t \rangle - \xi_t)^2 \|x_t\|^2$$

Telescope

$$0 \le \|w_T - w^*\|^2 = \|w_0 - w^*\|^2 - 4\eta \sum_t \langle w_{t-1} - w^*, x_t\rangle^2$$
$$+ 4\eta \sum_t \xi_t \langle x_t, w_{t-1} - w^* \rangle$$
$$+ 4\eta^2 \sum_t (\langle w_{t-1} - w^*, x_t\rangle - \xi_t)^2 \|x_t\|^2$$

Take expectation $\mathbb{E}\xi_t = 0$

"expected regret"

$$0 \le \|w_0 - w^*\|^2 - 4\eta \mathbb{E} \sum_t \langle w_{t-1} - w^*, x_t \rangle^2$$
$$+ 0 + 4\eta^2 \mathbb{E} \sum_t \langle w_{t-1} - w^*, x_t\rangle^2 \|x_t\|^2$$
$$+ 4\eta^2 \sum_t \mathbb{E} \xi_t^2 \|x_t\|^2$$

Suppose $\eta < \frac{1}{2R^2}$

$$2\eta \mathbb{E} \sum_t \langle w_{t-1} - w^*, x_t \rangle^2 \le \|w_0 - w^*\|^2 + 4\eta^2 R^2 \sum_t \mathbb{E} \xi_t^2$$

$$\mathbb{E} \sum_t \langle w_{t-1} - w^*, x_t \rangle^2 \le \frac{1}{2\eta} \|w_0 - w^*\|^2 + 2\eta R^2 \sum_t \mathbb{E} \xi_t^2$$

$$\simeq \|w_0 - w^*\| R \sqrt{\sum_t \xi_t^2}$$

for optimal choice of η.

ie "$O(\sqrt{T})$ ~~regret~~ rate"

Next

implies

~~Online~~ Convex Optimization

(Online Convex Optimization)

Suppose $F(x)$ is 1-strongly convex.

Then $\forall y, x$ $\quad F(x) - F(y) \le \langle \nabla F(x), x-y \rangle - \frac{1}{2}\|x-y\|^2$

$\Updownarrow$

$$F(y) \ge F(x) + \langle \nabla F(x), y-x \rangle + \frac{1}{2}\|y-x\|^2$$

$\Updownarrow$

$$F(y) - F(x) - \langle \nabla F(x), y-x \rangle \ge \frac{1}{2}\|y-x\|^2$$

(quadratic lower bound on F)

$F_0(x)$ 1-s.c.

$F_t(x) = F_{t-1}(x) + \ell_t(x)$ $\leftarrow$ linear $\langle g_t, x \rangle$ $F(y)$

Let $x_t = \operatorname{argmin} F_t(x)$

$\nabla F_t(x_{t-1}) = g_t$

$$\text{Regret} = \sum_{t=1} \ell_t(x_{t-1}) - \sum \ell_t(x^*)$$

$\uparrow$ oracle

Φ

$$\sum_{t=1} \langle g_t, x_{t-1} \rangle - \sum \langle g_t, x^* \rangle$$

$$= \sum_{t=1} \langle g_t, x_{t-1} - x^* \rangle$$

$F_t(x) = F_{t-1}(x) + \langle g_t, x \rangle$

$= \sum_{s=1}^{t} \langle g_s, x \rangle + \|x\|^2/2$

$\nabla F_t = 0 \rightarrow \sum g_s + x = 0$

$F_t($