$$f_i(\beta) = (y_i - x_i^T\beta)^2 \quad \text{(regression)}$$
$$f_i(\beta) = -y_i x_i^T \beta + \log(1 + \exp(x_i^T \beta))$$
(classification)

Lipschitz grad
strong convexity } grad descent uses fixed step sizes t

$$g_{i_k} = \nabla f_{i_k}(x^{(k-1)})$$

$$\mathbb{E}(\theta_\alpha) = \alpha\,\mathbb{E}(X) - \alpha\,\mathbb{E}(Y) + \mathbb{E}(Y)$$
$$= \alpha\,\mathbb{E}(X) + (1-\alpha)\,\mathbb{E}(Y)$$

$$\mathrm{Var}(\theta_\alpha) = \alpha^2\,\mathrm{Var}(X - Y)$$
$$= \alpha^2\left(\mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X,Y)\right)$$

$$f(x) = \mathbb{E}_\xi[F(x,\xi)]$$

$$x^{(k)} = x^{(k-1)} - t \, \nabla f(x^{(k-1)})$$

$$x^{(k)} = x^{(k-1)} - t H^{-1} \nabla f(x^{(k-1)})$$

$$H = \text{diag}^{1/2}\left(\sum g^{(l)} (g^{(l)})^T \right)$$

---

f convex.

GD bd : $\frac{R}{\sqrt{k}}$ $\quad R = \|x^{(0)} - x^*\|_2 \asymp \sqrt{p}.$

$$\sqrt{\frac{p}{k}}.$$

AdaGrad: $\sqrt{\frac{\log p}{k}}$ (sparse case)