Review: subgradient descent

- Initialize $x_1$
- for $t = 1$ to “I’m tired”:
  - $f_t =$ estimate of $f$
    - from limited # of terms (or just use $f$)
  - $g_t =$ any element of $\partial f_t(x_t)$
  - $x_{t+0.5} = x_t - \eta_t g_t$
  - $x_{t+1} = \arg\min_{x \in F} ||x - x_{t+0.5}||$
    - which is just $x_{t+1} = x_{t+0.5}$ if $F = \mathbb{R}^n$

Problem: $\min_x f(x)$ (optionally, s.t. $x \in F$)
Subgradient in action
Convergence summary

- For strictly convex $f(x)$ (i.e., $\lambda > 0$):
  - set $\eta_t = 1/\lambda t$
  - $f(x_t) - f(x^*) = \tilde{O}(1/t)$

- For non-strictly convex $f(x)$ (i.e., $\lambda = 0$):
  - set $\eta_t = 1/\sqrt{t}$
  - $f(x_t) - f(x^*) = O(1/\sqrt{t})$

- To get accuracy $\epsilon$:
  - $\lambda > 0$: $T = \tilde{O}(1/\epsilon)$
  - $\lambda = 0$: $T = O(1/\epsilon^2)$

Interior point: $T = O(\ln(1/\epsilon))$, but each iteration much slower
Convergence intuition

• $Q(x) = ||x - x^*||^2/2$

• Proof works by guaranteeing that $Q(x)$ decreases
  ‣ subtlety: only if $f(x_t) \gg f(x^*)$
  ‣ has to be like this: e.g., multiple minimizers of $f$

• We showed (for $\lambda=0$):
  ‣ $f(x^*) \geq f(x_t) + Q(x_{t+1})/\eta_t - Q(x_t)/\eta_t + \eta_t||g_t||^2/2$

• Suppose $f(x_t) \geq f(x^*) + \epsilon$:
Typical SVM bound

• Given n training examples, for any $\delta < 1$:
  ‣ define $E_{tr} =$
  ‣ w/ prob $\geq 1 - \delta$, test error rate $\leq$

• SVM optimizes $E_{tr}$ (s.t. $||w|| \leq$ const)
  ‣ let $E^* = \text{optimal } E_{tr}$
  ‣ suppose we solve to accuracy $\epsilon$:
    ‣ now suppose

• That is, accuracy needed is:
Example: SVM

• Suppose no b
  
  ‣ \( L = \|w\|^2/2 + (C/m) \sum h(y_ix_i^Tw) \)  
  
  ‣ \( \partial L = w + (C/m) \sum y_i x_i \partial h(y_ix_i^Tw) \)

• If \( \|x_i\| \leq X \):
What if we want b?

• Problem: $\lambda=0$

• Solutions:
  ‣ ignore the problem:
  ‣ penalize b too:
  ‣ change the algorithm slightly: