MATH 565: Lecture 3 (01/20/2026)

Today:
* more on SVM
* Taylor expansion
* local optimality conditions

Recall SVM

$$\min_{\bar{w}, b, \bar{\varepsilon}} \; J = C\sum_{i=1}^{n} \varepsilon_i + \frac{1}{2}\|\bar{w}\|^2 \qquad (C > 0)$$

$$\text{s.t.} \quad y_i(\bar{w}^T x_i + b) \geq 1 - \varepsilon_i, \quad i = 1, \dots, n$$

$$\varepsilon_i \geq 0 \;\; \forall i$$

Let's examine the unified (main) constraints in detail:

When $y_i = +1$, we get $\bar{w}^T x_i + b \geq 1 - \varepsilon_i$

ε_i : measures by how much the i^{th} sample violates well-separatedness

e.g., if $\bar{w}^T \bar{x}_i + b = 0.7$, then $\varepsilon_i \geq 0.3$

But if $\bar{w}^T \bar{x}_i + b = 2$, then $\varepsilon_i = 0$ works

When $y_i = -1$, we get $\bar{w}^T \bar{x}_i + b \leq -1 + \varepsilon_i$.

e.g., if $\bar{w}^T \bar{x}_i + b = -3$, $\varepsilon_i = 0$ in the opt. soln.

But if $\bar{w}^T \bar{x}_i + b = 0.5$, we need $\varepsilon_i \geq 1.5$

Can also regularize b (intercept term).

Or, take $\bar{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}$ as $(d+1)$-variable vector, and

write $J = \min_{\bar{w}, \bar{\varepsilon}} \; C\sum \varepsilon_i + \frac{1}{2}\|\bar{w}\|^2$

$$\text{s.t.} \quad y_i(\bar{w}_{[1:d]}^T \bar{x}_i + w_0) \geq 1 - \varepsilon_i, \quad i = 1, \dots, n$$

$$\varepsilon_i \geq 0 \;\; \forall i.$$

Taylor Expansion

In 1D, Taylor expansion of $f(x)$ at $x=a$ is

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a) + \dots + \frac{(x-a)^r}{r!} \frac{d^r f(x)}{dx^r}\bigg|_{x=a} + \dots$$

When $|x-a|$ is small, we could take

$$f(x) \approx f(a) + (x-a)f'(a) + \frac{(x-a)^2}{2!} f''(a)$$

In d-dimensions, Taylor expansion of $f(\bar{x})$ at $\bar{x}=\bar{a}$

$$f(\bar{x}) = f(\bar{a}) + \sum_{i=1}^{d} (x_i - a_i)\left[\frac{\partial f}{\partial x_i}\right]\bigg|_{\bar{x}=\bar{a}} + \sum_i \sum_j \frac{(x_i-a_i)(x_j-a_j)}{2!}\left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]\bigg|_{\bar{x}=\bar{a}}$$

$$+ \dots$$

Equivalently,

$$f(\bar{x}) = f(\bar{a}) + [\bar{x}-\bar{a}]^T \nabla f(\bar{a}) + [\bar{x}-\bar{a}]^T Hf(a)[\bar{x}-\bar{a}] + \dots$$

Local Optimality Conditions (in 1D)

Lemma 1 $f(x)$ is a minimum value at $x = x_0$ with respect to its immediate locality if → sufficient condition

$$\underbrace{f'(x_0) = 0}_{\text{first order condition for optimality}} \quad \text{and} \quad \underbrace{f''(x_0) > 0}_{\text{second order condition for optimality}}.$$

Proof Consider Taylor expansion of $f(\cdot)$ at x_0 $(x = x_0 + \Delta)$

$$f(x_0+\Delta) \approx f(x_0) + \Delta \underbrace{f'(x_0)}_{=0} + \frac{\Delta^2}{2} \underbrace{f''(x_0)}_{>0} \quad \text{for } |\Delta| << 1.$$

small enough

$\Rightarrow$ under first and second order optimality conditions,

$$f(x_0+\Delta) > f(x_0) \quad \text{for } |\Delta| \text{ small enough.}$$

□

The first order condition $(f'(x) = 0)$ is typically solved using gradient descent.

Step 0. Start $x = x_0$ (randomly chosen)

Step k. $x_k \leftarrow x_{k-1} - \alpha f'(x_{k-1})$

(in general, $x \leftarrow x - \alpha f'(x)$)

α: learning rate (step size)

Changing x by $\delta x = -\alpha f'(x)$.

We're changing x along the "steepest descent" direction (trivial in 1D, nontrivial in d-dim.).

This change will reduce $f(x)$ for "small" values of α:

$$f(x+\delta x) \approx f(x) + \delta x\, f'(x)$$
$$= f(x) - \alpha\left[f'(x)\right]^2$$
$$< f(x)$$

Example

$$f(x) = x^2 \sin(x) + x$$
$$f'(x) = 2x\sin(x) + x^2\cos(x) + 1 \stackrel{?}{=} 0$$
$$f''(x) = (2-x^2)\sin(x) + 4x\cos(x)$$

Local Optimality in d-dim

Variables: $\bar{w}$ $(w_1, \dots, w_d, w_0)$, $\xi_i, \dots$

Obj. fn: loss function J e.g., $J = \frac{1}{2}\|D\bar{w} - \bar{y}\|^2 + \frac{1}{2}\|\bar{w}\|^2$

1st order condition: $\nabla J = \bar{0}$

$$\begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \frac{\partial J}{\partial w_d} \end{bmatrix} = \bar{0}$$

2nd order condition: $H \succ 0$ Hessian is positive definite

i.e., $\bar{x}^T H \bar{x} > 0 \;\; \forall \bar{x} \in \mathbb{R}^d / \{\bar{0}\}$

Taylor expansion: $\epsilon > 0$

$$J(\bar{w}_0 + \epsilon \bar{v}) \simeq J(\bar{w}_0) + \epsilon \bar{v}^T \underbrace{\nabla J(\bar{w}_0)}_{=0} + \frac{\epsilon^2}{2} \underbrace{\bar{v}^T H \bar{v}}_{>0 \;\; \forall \bar{v}}$$