PC: Feature Transform

Song Liu

Polynomial Feature Transform

- Given i, b and d, The i-th dimension of polynomial feature transform $\boldsymbol{f}(\boldsymbol{x})$ has the form $f^{(i)}(\boldsymbol{x}) = \left(x^{(m)}\right)^n$. Express m and n using i, b and d.

- $m = \left\lfloor\frac{i-1}{b+1}\right\rfloor + 1, n = (i-1)\%(b+1)$
  - $\left\lfloor\frac{a}{b}\right\rfloor, a\%b$ are quotient and remainder.

Computational Complexity

- Computational complexity of the LS solution of $\widehat{\boldsymbol{\beta}}$ using $\boldsymbol{f}(\boldsymbol{x}) \in R^m$?
- $\widehat{\boldsymbol{\beta}} = \Big(\underbrace{\boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{f}(\boldsymbol{X})}_{A}\Big)^{-\mathbf{1}} \underbrace{\boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{y}}_{B}$
  - $\boldsymbol{A} = \boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{f}(\boldsymbol{X})$: $O(m \times n \times m)$
  - $(\boldsymbol{A})^{-\mathbf{1}}$: $O(m^3)$
  - $\boldsymbol{B} = \boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{y}$: $O(m \times n \times 1)$
  - $(\boldsymbol{A})^{-1}\boldsymbol{B}$: $O(m \times m \times 1)$
  - **Fixing n, comp. complexity is $O(m^3)$.**

Expected Error

- Rewrite $\mathbb{E}_\epsilon[(y - \hat{y}_i)^2 | \boldsymbol{x}_i]$ using integral.
  - Conditional Expectation:
  - $\mathbb{E}[g(y)|x_i] = \int p(y|x) g(y) dy$.

  - $$\mathbb{E}_\epsilon[(y - \hat{y})^2 | \boldsymbol{x}_i] = \int p(y|\boldsymbol{x}_i)\,(y - \hat{y})^2 dy$$
    $$= \int p(y|\boldsymbol{x}_i)\left(y - \langle \widehat{\boldsymbol{\beta}}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle\right)^2 dy$$
    $$= \int p(y|\boldsymbol{x}_i)\,(y - \langle (\boldsymbol{F}^\top \boldsymbol{F})^{-1} \boldsymbol{F}^\top \boldsymbol{y}, \boldsymbol{f}(\boldsymbol{x}_i)\rangle)^2 dy$$
  - Note the difference between y and $\boldsymbol{y}$.

Expanding Variance Term

- Show $\mathrm{var}[\hat{y}|\boldsymbol{x}_i]$ is $< h(\boldsymbol{x}_i), h(\boldsymbol{x}_i) > \cdot\ \sigma^2$
  - $h(\boldsymbol{x}_i) \coloneqq \boldsymbol{f}(\boldsymbol{x}_i)\left(\boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{f}(\boldsymbol{X})\right)^{-1} \boldsymbol{f}(\boldsymbol{X})^\top$
- **Hint:**
  - $y = \boldsymbol{f}(\boldsymbol{x}_i) + \epsilon$
  - Shorten $\boldsymbol{f}(\boldsymbol{x})$ as $\boldsymbol{f}$, $\boldsymbol{f}(\boldsymbol{X})$ as $\boldsymbol{F}$, $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2 \ldots \epsilon_n\ ]^\top$.
  - $\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{x}A\boldsymbol{x}^\top] = \mathrm{tr}[A\Sigma]$, Σ is the covar. of $\boldsymbol{x}$.

  - Live demonstration

Induced Kernel

- Write down induced $\boldsymbol{f}(\boldsymbol{x})$ by polynomial kernels b=2.
- $k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left(< \boldsymbol{x}_i, \boldsymbol{x}_j > +1\right)^b$
- **Hint:**
  - Express k using inner product of two vec.

$$k\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right) = \left[f^{(1)}(\boldsymbol{x}) \ldots f^{(m)}(\boldsymbol{x})\right] \cdot \begin{bmatrix} f^{(1)}(\boldsymbol{x}) \\ \cdots \\ f^{(m)}(\boldsymbol{x}) \end{bmatrix}$$

  - $\boldsymbol{f}(\boldsymbol{x}) = \left[1, x^{(1)} \cdots x^{(d)}, \left(x^{(1)}\right)^2 \cdots \left(x^{(d)}\right)^2, \forall_{u \neq v} x^{(u)} x^{(v)}\right]$