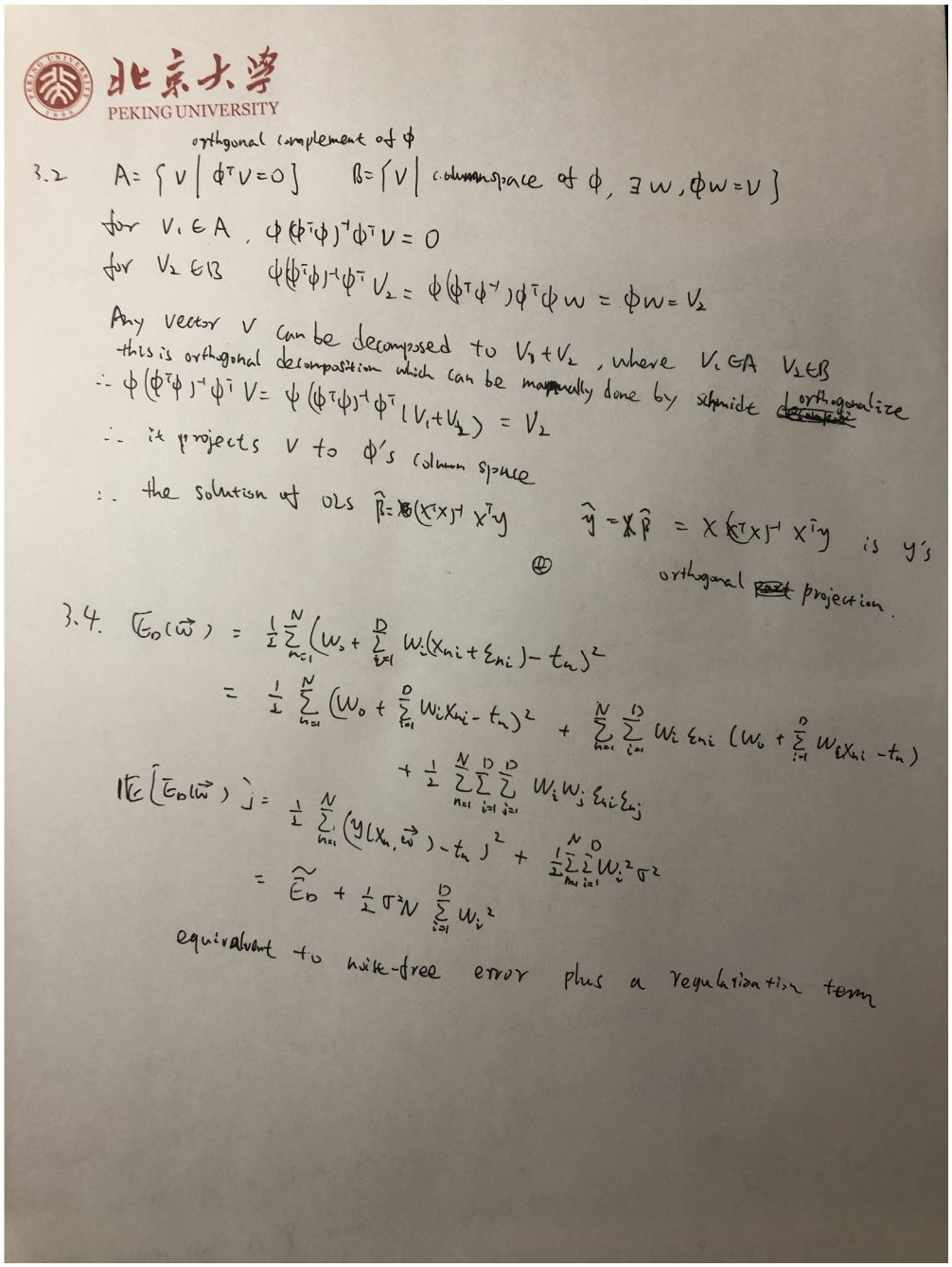
HW1

MTH9898_Chenyu_Zhao



HW1

3.5

① $\frac{1}{2}\sum_{i=1}^{N}[t_n - w^T\phi(x_n)]^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$

② $\frac{1}{2}\sum_{i=1}^{N}[t_n - w^T\phi(x_n)]^2$ s.t. $\sum_{j=1}^{M}|w_j|^q \le \eta$

②: $L = \frac{1}{2}\sum_{i=1}^{N}[t_n - w^T\phi(x_n)]^2 + \frac{\lambda'}{2}\left(\sum_{j=1}^{M}|w_j|^q - \eta\right)$ λ' is langrange multiplier, $\lambda' \ge 0$

it has same expression with ①

when $\lambda \nearrow$, $\eta \searrow$

$\eta = \sum_{i=1}^{M}|w_i^*(\lambda)|^q$

2.3 ~~$P(d > x) = \left(\frac{x^p}{1}\right)^N$~~

d = distance from origin to cloest data point

one point's possibility that distance bigger than x is $1-x^p$, ∵ N points

$\therefore P(d > x) = (1-x^p)^N$

$P(d \le x) = 1 - (1-x^p)^N = \frac{1}{2}$

$x = \left(1 - \frac{1}{2}^{\frac{1}{N}}\right)^{\frac{1}{p}}$

2.9 $\mathbb{E}\left[\frac{1}{m}\sum_{i=1}^{m}(\tilde{y}_i - \hat{\beta}^T\tilde{x}_i)\right] = \mathbb{E}(\tilde{y}_1 - \hat{\beta}^T\tilde{x}_1)$ because $(\tilde{x}_i, \tilde{y}_i)$ are i.i.d.

$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_i - \hat{\beta}^T\tilde{x}_i)\right] := R_{te}'(\hat{\beta})$

now we compare $\mathbb{E}R_{tr}(\hat{\beta})$ and $\mathbb{E}R_{te}'(\hat{\beta})$ where they have same size n

$\mathbb{E}R_{tr}(\hat{\beta}) = \mathbb{E}R_{te}'(\tilde{\beta})$ where $\tilde{\beta}$ is trained on test set. because the two sets have same distibution

And $\mathbb{E}R_{te}'(\tilde{\beta}) \le \mathbb{E}R_{te}'(\hat{\beta}) = \mathbb{E}R_{te}(\hat{\beta})$, because $\tilde{\beta}$ is least square fitted

$\therefore \mathbb{E}R_{tr}(\hat{\beta}) \le \mathbb{E}R_{te}(\hat{\beta})$