



# 核函数低维随机特征逼近的广义构建方法 及应用研究

汇报人：罗钦

专    业：控制工程

导  师：黄晓霖 副教授

申请学位：硕士

2022年02月15日



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY



## 1 研究背景与意义

## 2 不定核随机特征构建方法研究

## 3 低比特量化随机特征的构建研究

## 4 基于神经正切核低维假设的小样本学习研究

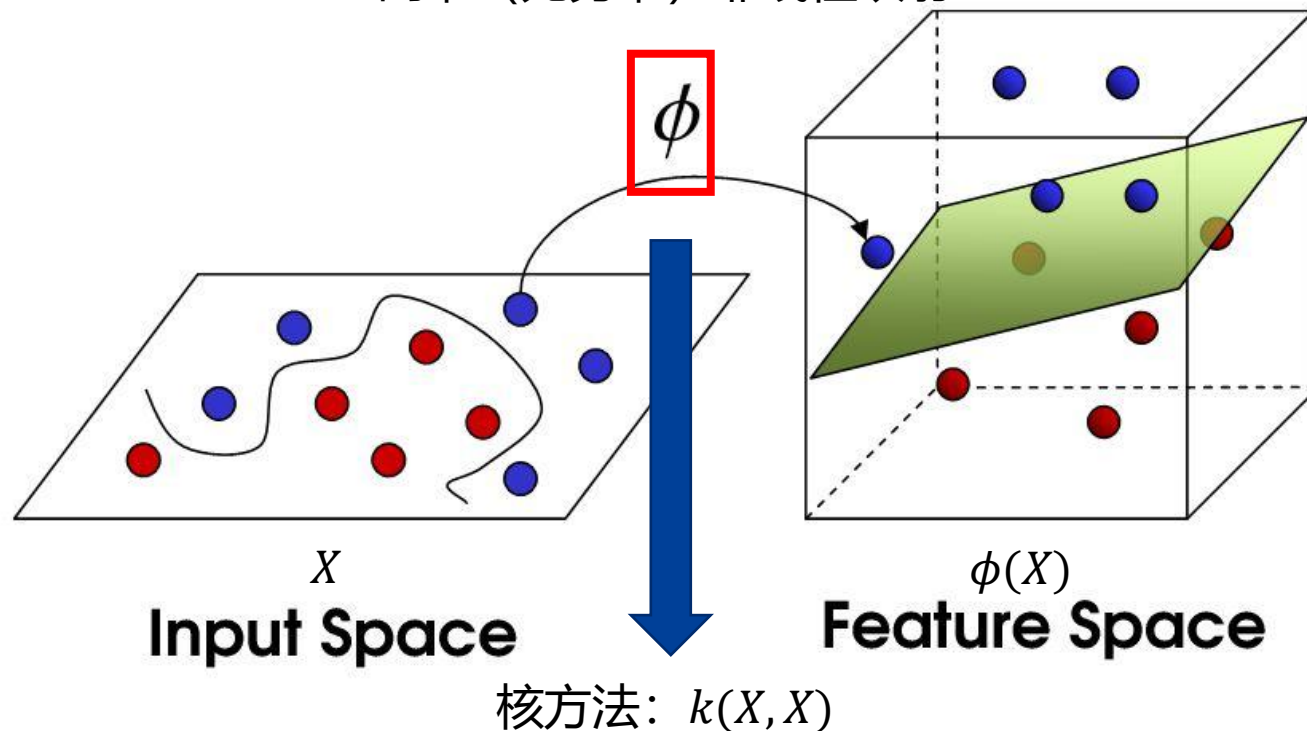
## 5 结论与展望

# 研究背景与意义



核方法：利用样本之间的相似度关系来代替高维（无穷维）的非线性映射

高维（无穷维）非线性映射



特点：表示能力强、易于训练、坚实的理论保证

应用：分类（SVM, KLR）、回归（SVR, GPR）、降维（kPCA）、聚类（KSC）



## 随机特征逼近的背景

### 传统核方法存在的问题

- **大规模核学习问题中运算和存储复杂度较高**
  - 运算复杂度 $\mathcal{O}(N^2 d)$ , 存储复杂度 $\mathcal{O}(N^2)$ , **与N成平方关系**
- **核函数选择范围有限**
  - 受限于经验定义核函数的形式, 依赖于先验知识

$$k(x, y) \approx \langle \phi(x), \phi(y) \rangle$$



$\phi(x)$ 称为**随机特征**, 通过**较低的维度**对于**无穷维**算子进行无偏逼近

- **大规模核学习问题**
  - 运算复杂度 $\mathcal{O}(Ns^2)$ , 存储复杂度 $\mathcal{O}(Ns)$ , **与N成线性关系**
- **灵活的核学习问题**
  - 覆盖整个**正定核函数**空间, 经验驱动→**数据驱动**

# 研究背景与意义

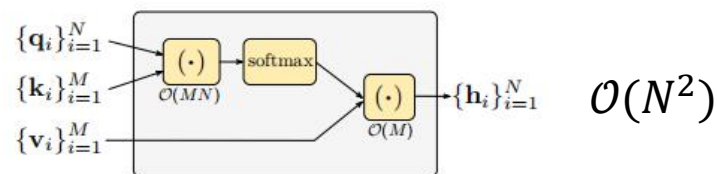


## 随机特征逼近的应用背景

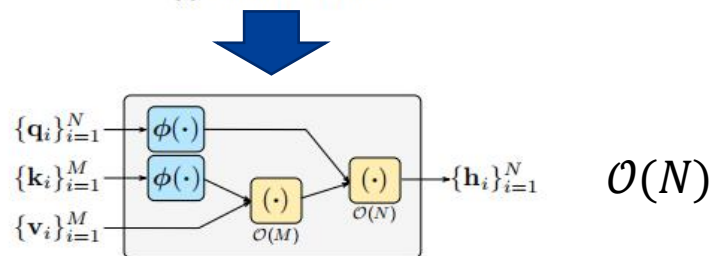
### 机器学习算法压缩加速

1. 大规模核学习算法加速
2. 基于相似度（不相似度）的机器学习算法的加速

e.g. Transformer加速



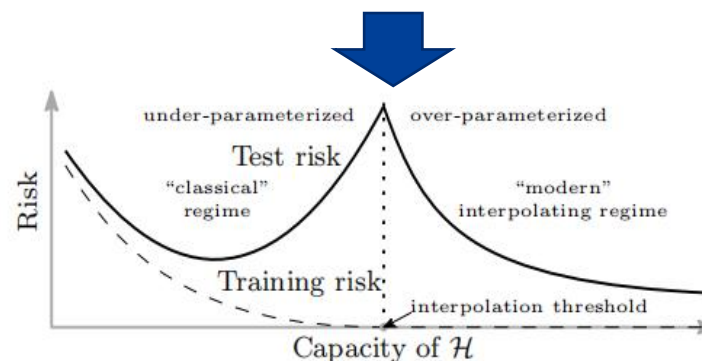
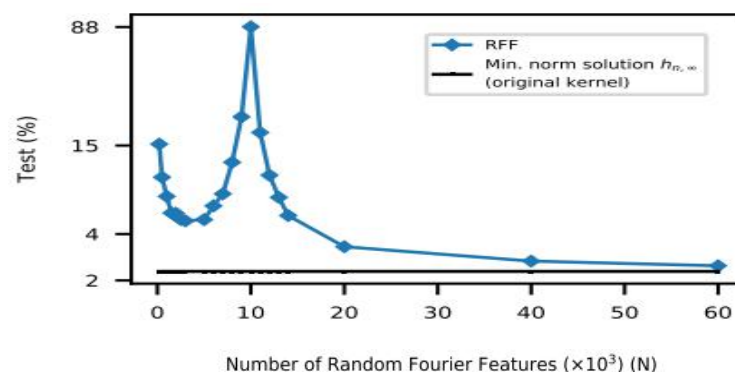
(a) Softmax attention.



(b) Random feature attention.

### 解释神经网络动态和泛化性能

#### Double Descent现象



[1] Choromanski K M, Likhoshesterov V, Dohan D, et al. Rethinking Attention with Performers[C]//ICLR 2020.

[2] Belkin M, Hsu D, Xu J. Two models of double descent for weak features[J]. SIAM Journal on Mathematics of Data Science, 2020, 2(4): 1167-1180.



## 随机特征逼近构建方法的国内外研究现状

### 与本论文相关的随机特征构建方法研究综述

#### Data Independent Random Features

**RFF**(2007, test-of-time award on NeurIPS2017)  
Structral Matrix: **Fastfood**(2013), **SCRF**(2015),  
p-model(2015)  
Orthogonality: **ORF**, **SORF**(2016)  
Quasi Monte Carlo: **QMC**(2014), **MM**, **SSF**(2017)

#### Data Dependent Random Features

leverage score: **LS-RFF**(2019), **SLS-RFF**(2020)  
kernel alignment: **EE-RFF**(2018)  
end-to-end learning: **DKR**(2019), **Generative Method**(2020)

#### 存在的问题:

- 1) 现有随机特征逼近框架局限于**正定性和平移不变核函数**
- 2) **低计算存储资源下基于随机特征的核方法的泛化能力下降**  
(计算存储资源有限→过少的随机特征数目→核方法泛化能力下降)

[1] Rahimi A, Recht B. Random features for large-scale kernel machines. NIPS 2007.

[2] Liu F, Huang X, et al. Random Features for Kernel Approximation: A Survey on Algorithms, Theory, and Beyond. TPAMI 2021.

[3] Yu F X X, Suresh A T, et al. Orthogonal random features. NIPS 2016.

[4] Rudi A, Calandriello D, et al. On fast leverage score sampling and optimal learning. NIPS 2018.

[5] Fang K, Huang X, et al. End-to-end Kernel Learning via Generative Random Fourier Features. arXiv preprint 2020.



## 本论文主要研究的内容及意义

随机特征方法的  
适用范围?

随机特征如何更  
好地构建? (方差  
减小、泛化性能  
提升等)

随机特征的应用?

不定核低维随机特征构建方法  
研究

**(突破现有随机特征框架  
正定性和平移不变性的约束)**

低比特量化随机特征的构建  
研究

**(在计算存储资源有限情况下  
构建更好的随机特征逼近方式)**

基于NTK低维假设的小样  
本学习研究

**(从核函数低维逼近出发  
构造深度神经网络在小样  
本学习问题中的低维优化  
空间)**

构建方法的理论研究

应用研究



## 1 研究背景与意义

## 2 不定核随机特征构建方法研究

## 3 低比特量化随机特征的构建研究

## 4 基于神经正切核低维假设的小样本学习研究

## 5 结论与展望



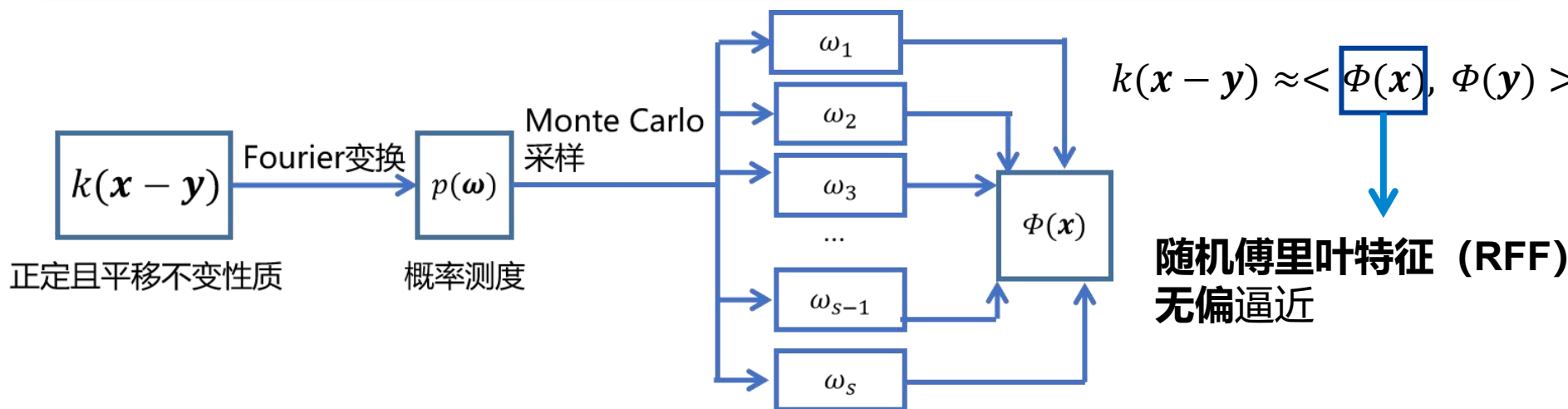
## 随机傅里叶特征基本理论

### Bohner定理

定理 2.1. (Bochner 定理<sup>[99]</sup>) 定义连续且平移不变核函数  $k(.,.): \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ , 核函数  $k(.,.)$  是正定核函数的充要条件是其可以表示成如下形式:

$$k(x-y) = \int_{\mathbb{R}^d} \exp(i\omega^\top(x-y)) p(d\omega) \quad (2.2)$$

其中  $p(\omega)$  是定义在频率  $\omega$  上的非负且有限的 Borel 测度。



$$\Phi(x) = \frac{1}{\sqrt{s}} [\cos(\omega_1^T x), \dots, \cos(\omega_s^T x), \sin(\omega_s^T x), \dots, \sin(\omega_s^T x)]^T$$



## 不定核随机特征的研究意义

### 1) 不定核函数的广泛应用和良好的效果

- 实际应用中的采用部分核函数为不定核，如Epanechnikov核函数（多项式核、线性核采用随机特征），衡量蛋白质序列相似度的核函数。
- 一些不定核函数在某些任务上具有比正定核函数更好的性能，如Huang等人提出的TL1核函数在C-SVM分类任务上具有比RBF核函数更好的性能。

### 2) 根据正定核函数定义判断核函数正定操作困难

### 3) 现有不定核逼近方法的逼近误差大

| 方法                               | 无偏性 | 逼近方差                                                                                                                  |
|----------------------------------|-----|-----------------------------------------------------------------------------------------------------------------------|
| Random Maclaurin                 | √   | $\mathcal{O}\left(\left(\frac{32RL}{\epsilon}\right)^{2d} \exp\left(-\frac{D\epsilon^2}{8C_{\Omega}^2}\right)\right)$ |
| Tensor Sketch                    | √   | $\mathcal{O}\left(\exp\left(-\frac{t\epsilon^2}{2R^{4p}}\right)\right)$                                               |
| Spherical Random Features        | ×   | _____                                                                                                                 |
| Double Variation Random Features | ×   | _____                                                                                                                 |

[1] Kar P, Karnick H. Random feature maps for dot product kernels AISTATS 2012.

[2] Pham N, Pagh R. Fast and scalable polynomial kernels via explicit feature maps SIGKDD 2013.

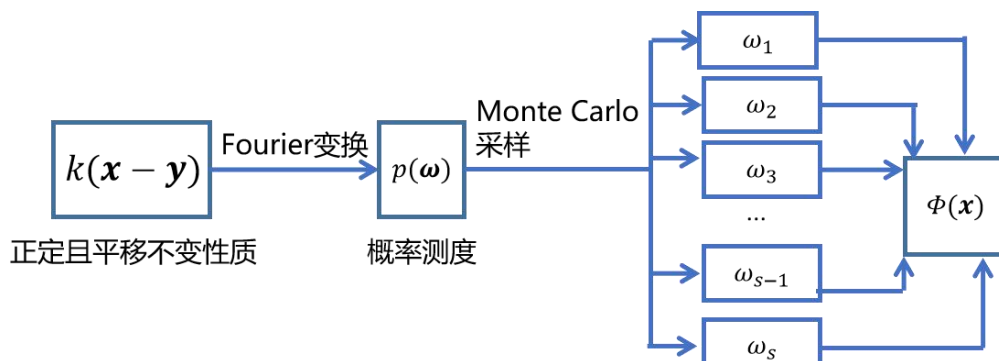
[3] Pennington J, Yu F X X, et al. Spherical random features for polynomial kernels. NIPS 2015.

[4] Liu F, Huang X, Shi L, et al. A double-variational bayesian framework in random fourier features for indefinite kernels. TNNLS 2019.

# 不定核随机特征构建方法研究



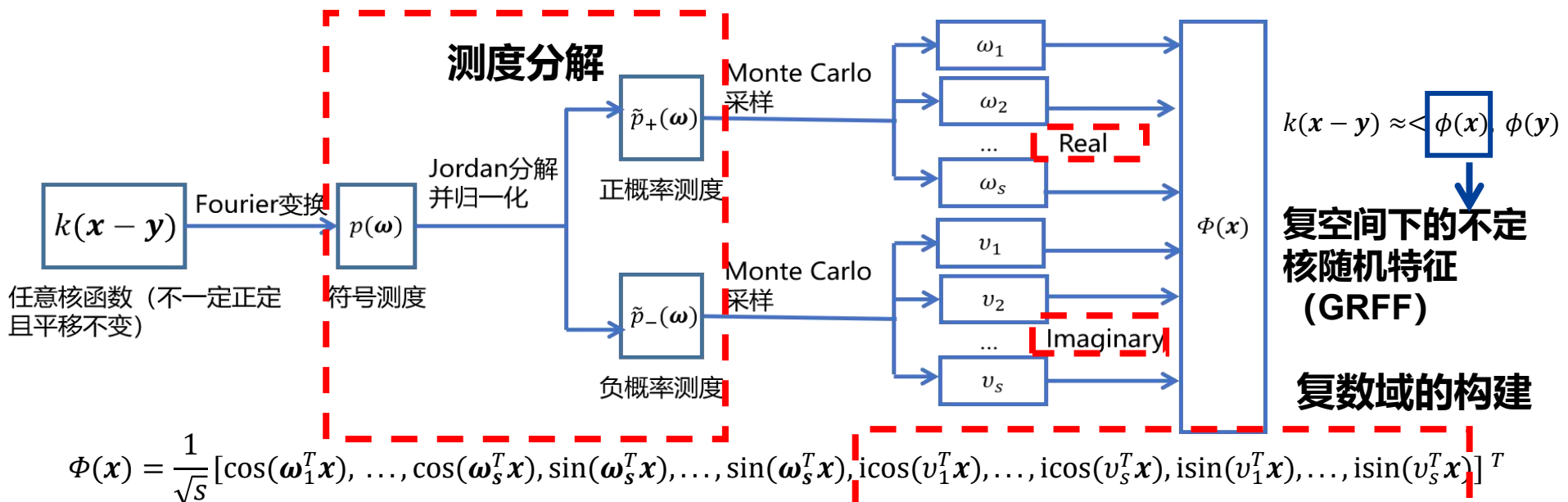
## 复空间下的不定核随机特征构建方法



$$k(x-y) \approx \langle \phi(x), \phi(y) \rangle$$

正定核的随机特征 (RFF)

$$\Phi(x) = \frac{1}{\sqrt{s}} [\cos(\omega_1^T x), \dots, \cos(\omega_s^T x), \sin(\omega_s^T x), \dots, \sin(\omega_s^T x)]^T$$



$$\Phi(x) = \frac{1}{\sqrt{s}} [\cos(\omega_1^T x), \dots, \cos(\omega_s^T x), \sin(\omega_s^T x), \dots, \sin(\omega_s^T x), i\cos(v_1^T x), \dots, i\cos(v_s^T x), i\sin(v_1^T x), \dots, i\sin(v_s^T x)]^T$$



## 逼近方差进一步减小的方法

复空间下的不定核随机特征 (GRFF) 是**无偏逼近**, 逼近方差决定随机特征的逼近误差。  
**正交化 (Orthogonality)**

### 算法 1 GORF 构建方法

输入: 平移不变核函数  $k(x, y) = k(z)$ ,  $z = \|x - y\|$ , 训练样本的维度  $d$ , GORF 的数目  $s$

输出: 满足  $k(x - y) \approx \phi(x)^T \phi(y)$  的 GORF 映射  $\phi(x)$

- 1: 通过核函数的随机傅里叶变换获取符号测度  $p(\omega)$ , 并且通过 Jordan 分解与归一化计算  $p_+(\omega)$ ,  $p_-(v)$ ,  $\|p_+\|$ ,  $\|p_-\|$ ,  $\tilde{p}_+(\omega)$ ,  $\tilde{p}_-(v)$ 。
- 2: 分别从  $\tilde{p}_+(\omega)$  和  $\tilde{p}_-(v)$  中采样  $s$  个  $\ell_2$  范数的权值的模, 即  $\|\omega_i\|_2$  和  $\|v_i\|_2$ 。
- 3: 从标准正态分布  $N(0, I_{2m})$  采样  $2m$  方向向量  $a_j$  和  $b_j$ , 并组合成矩阵  $M$ 。利用 QR 分解获得正交矩阵  $M^{orth}$  和正交单位矩阵  $M^{orthn}$ 。
- 4: 根据公式(2.34) 得到  $\omega_i$  和  $v_i$ 。
- 5: 根据公式(2.24)和(2.25)产生  $\psi_i(x)$  和  $\phi(x)$ 。

→ GRFF构建

→ 正交化

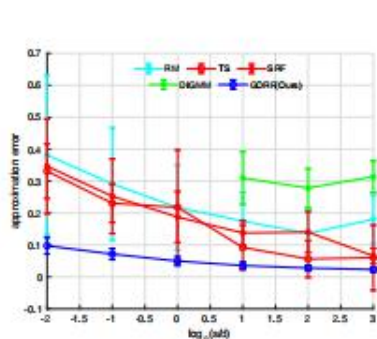
# 不定核随机特征构建方法研究



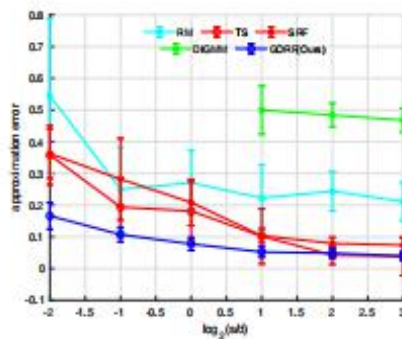
## 实验-不同不定核逼近方法的逼近误差的对比

$$\text{逼近误差} = \frac{\|K - \hat{K}\|_F}{\|K\|_F}$$

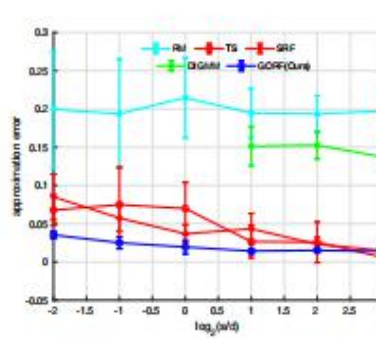
### 不同不定核逼近方法在不同核函数及数据集上的逼近误差对比



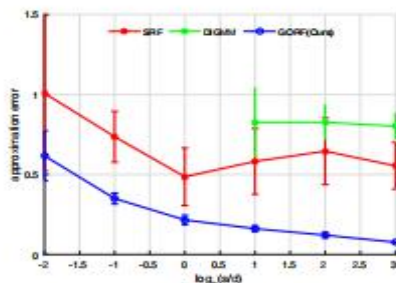
(a) letter 数据集, Epanechnikov 核函数 ( $a=3, m=1$ )



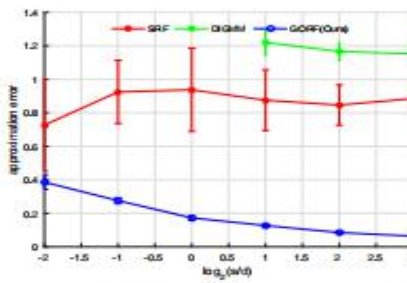
(c) ijcnn1 数据集, Epanechnikov 核函数 ( $a=3, m=1$ )



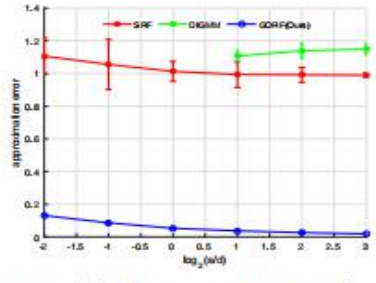
(e) usps 数据集, Epanechnikov 核函数 ( $a=3, m=1$ )



(b) letter 数据集, Delta-Gaussian 核函数 ( $a_1=1, a_2=-1, \sigma_1=1, \sigma_2=10$ )



(d) ijcnn1 数据集, Delta-Gaussian 核函数 ( $a_1=1, a_2=-1, \sigma_1=1, \sigma_2=10$ )



(f) usps 数据集, Delta-Gaussian 核函数 ( $a_1=1, a_2=-1, \sigma_1=1, \sigma_2=10$ )

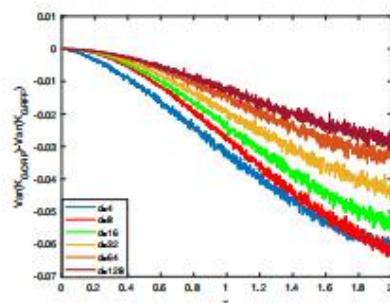
深蓝线——  
GORF方法  
1) 无偏性  
2) 更小的方差 (RFF和正交化)



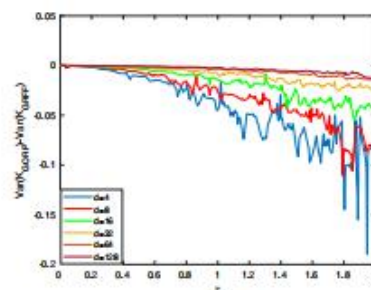
# 不定核随机特征构建方法研究



## 实验-正交化对不定核随机特征的逼近误差的影响



(a) Epanechnikov 核函数 ( $a = 3, m = 1$ )



(b) Delta-Gaussian 核函数 ( $a_1 = 1, a_2 = -1, \sigma_1 = 1, \sigma_2 = 10$ )

理论计算  
 $\mathbb{V}[GORF] - \mathbb{V}[GRFF] < 0$

### 正交化前后不定核随机特征逼近的逼近误差实验对比

| 核函数            | 数据集   | 方法   | $s = 1/2d$                            | $s = d$                               | $s = 2d$                              |
|----------------|-------|------|---------------------------------------|---------------------------------------|---------------------------------------|
| Epanechnikov   | lette | GRFF | $0.0859 \pm 0.0309$                   | $0.0547 \pm 0.0078$                   | $0.0469 \pm 0.0109$                   |
|                |       | GORF | <b><math>0.0716 \pm 0.0175</math></b> | <b><math>0.0495 \pm 0.0139</math></b> | <b><math>0.0360 \pm 0.0110</math></b> |
|                | ijcnn | GRFF | $0.1159 \pm 0.0158$                   | $0.0907 \pm 0.0194$                   | $0.0794 \pm 0.0142$                   |
|                |       | GORF | <b><math>0.1072 \pm 0.0228</math></b> | <b><math>0.0775 \pm 0.0204</math></b> | <b><math>0.0487 \pm 0.0155</math></b> |
|                | usps  | GRFF | $0.0270 \pm 0.0056$                   | $0.0213 \pm 0.0063$                   | $0.0160 \pm 0.0029$                   |
|                |       | GORF | <b><math>0.0251 \pm 0.0078</math></b> | <b><math>0.0194 \pm 0.0087</math></b> | <b><math>0.0143 \pm 0.0030</math></b> |
| Delta-Gaussian | lette | GRFF | $0.3918 \pm 0.0428$                   | $0.2736 \pm 0.0345$                   | $0.1887 \pm 0.0201$                   |
|                |       | GORF | <b><math>0.3154 \pm 0.0424</math></b> | <b><math>0.1133 \pm 0.0181</math></b> | <b><math>0.0760 \pm 0.0090</math></b> |
|                | ijcnn | GRFF | $0.2924 \pm 0.0188$                   | $0.2171 \pm 0.0222$                   | $0.1504 \pm 0.0134$                   |
|                |       | GORF | <b><math>0.2415 \pm 0.0190</math></b> | <b><math>0.1026 \pm 0.0129</math></b> | <b><math>0.0739 \pm 0.0065</math></b> |
|                | usps  | GRFF | $0.1005 \pm 0.0061$                   | $0.0690 \pm 0.0050$                   | $0.0500 \pm 0.0024$                   |
|                |       | GORF | <b><math>0.0724 \pm 0.0049</math></b> | <b><math>0.0235 \pm 0.0009</math></b> | <b><math>0.0166 \pm 0.0008</math></b> |

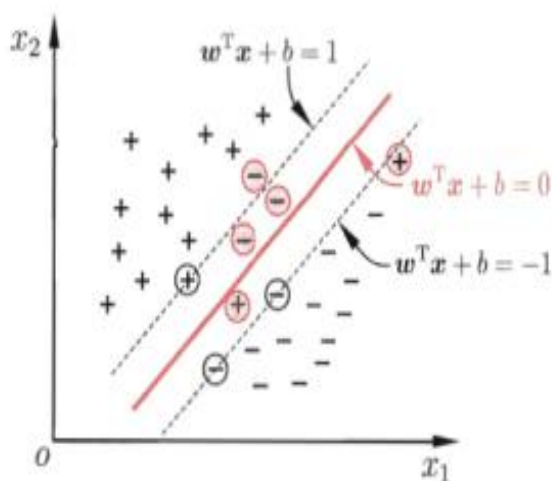


# 不定核随机特征构建方法研究

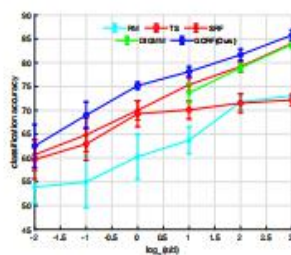


## 实验-分类问题上不同不定核逼近方法对比

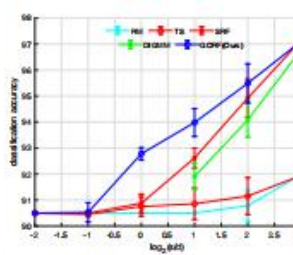
### SVM分类问题上的实验结果 (分类准确率)



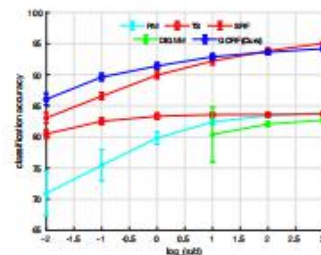
c-SVM问题



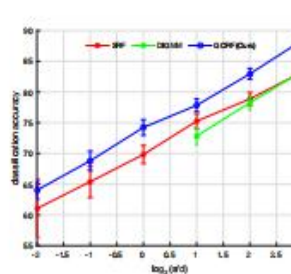
(a) letter 数据集, Epanechnikov 核函数 ( $a = 3, m = 1$ )



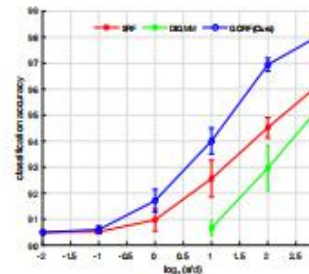
(c) ijcm1 数据集, Epanechnikov 核函数 ( $a = 3, m = 1$ )



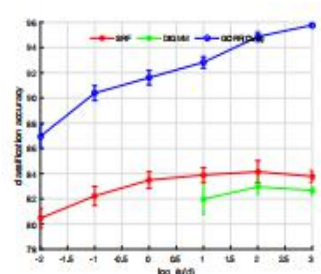
(e) usps 数据集, Epanechnikov 核函数 ( $a = 3, m = 1$ )



(b) letter 数据集, Delta-Gaussian 核函数 ( $a_1 = 1, a_2 = -1, \sigma_1 = 1, \sigma_2 = 10$ )



(d) ijcm1 数据集, Delta-Gaussian 核函数 ( $a_1 = 1, a_2 = -1, \sigma_1 = 1, \sigma_2 = 10$ )



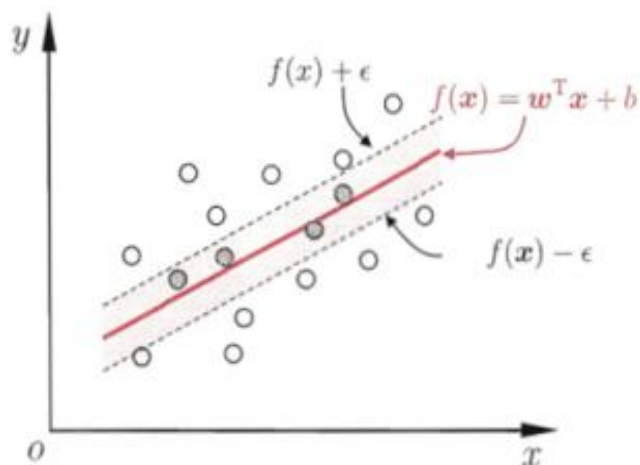
(f) usps 数据集, Delta-Gaussian 核函数 ( $a_1 = 1, a_2 = -1, \sigma_1 = 1, \sigma_2 = 10$ )

深蓝线——GORF方法 逼近误差↓→分类准确率↑

# 不定核随机特征构建方法研究



## 实验-回归问题上不同不定核逼近方法对比



$\epsilon$ -SVR问题

### SVR回归问题上的实验结果 (RMSE)

| 核函数            | 数据集     | 方法         | $s = 2d$                            | $s = 4d$                            | $s = 8d$                            |
|----------------|---------|------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Epanechnikov   | mpg     | RM         | $7.137 \pm 1.828$                   | $5.623 \pm 0.862$                   | $4.707 \pm 0.107$                   |
|                |         | TS         | $5.050 \pm 0.759$                   | $4.952 \pm 0.423$                   | $4.674 \pm 0.117$                   |
|                |         | SRF        | $4.461 \pm 0.136$                   | $4.269 \pm 0.129$                   | $4.137 \pm 0.081$                   |
|                | housing | DIGMM      | $4.686 \pm 0.287$                   | $4.339 \pm 0.128$                   | $4.133 \pm 0.119$                   |
|                |         | GORF(OURS) | <b><math>4.342 \pm 0.102</math></b> | <b><math>4.162 \pm 0.157</math></b> | <b><math>3.872 \pm 0.117</math></b> |
|                |         | RM         | $7.153 \pm 1.772$                   | $5.436 \pm 0.917$                   | $4.491 \pm 0.008$                   |
| Delta-Gaussian | mpg     | TS         | $5.414 \pm 0.879$                   | $4.772 \pm 0.377$                   | $4.657 \pm 0.316$                   |
|                |         | SRF        | $4.391 \pm 0.368$                   | $3.906 \pm 0.219$                   | $3.555 \pm 0.130$                   |
|                |         | DIGMM      | $4.897 \pm 0.368$                   | $4.130 \pm 0.324$                   | $4.000 \pm 0.280$                   |
|                | housing | GORF(OURS) | <b><math>4.079 \pm 0.233</math></b> | <b><math>3.817 \pm 0.204</math></b> | <b><math>3.472 \pm 0.137</math></b> |
|                |         | SRF        | $5.243 \pm 0.110$                   | $5.189 \pm 0.095$                   | $4.958 \pm 0.090$                   |
|                |         | DIGMM      | $5.203 \pm 0.212$                   | $4.925 \pm 0.215$                   | $4.613 \pm 0.111$                   |
| Delta-Gaussian | mpg     | GORF(OURS) | <b><math>4.831 \pm 0.173</math></b> | <b><math>4.622 \pm 0.189</math></b> | <b><math>4.488 \pm 0.128</math></b> |
|                |         | SRF        | $5.432 \pm 0.729$                   | $3.845 \pm 0.379$                   | $3.321 \pm 0.274$                   |
|                | housing | DIGMM      | $4.647 \pm 0.411$                   | $3.898 \pm 0.598$                   | $3.688 \pm 0.192$                   |
|                |         | GORF(OURS) | <b><math>3.739 \pm 0.360</math></b> | <b><math>3.474 \pm 0.330</math></b> | <b><math>3.164 \pm 0.252</math></b> |

GORF方法 逼近误差 $\downarrow \rightarrow$ RMSE $\downarrow$



## 1 研究背景与意义

## 2 不定核随机特征构建方法研究

## 3 低比特量化随机特征的构建研究

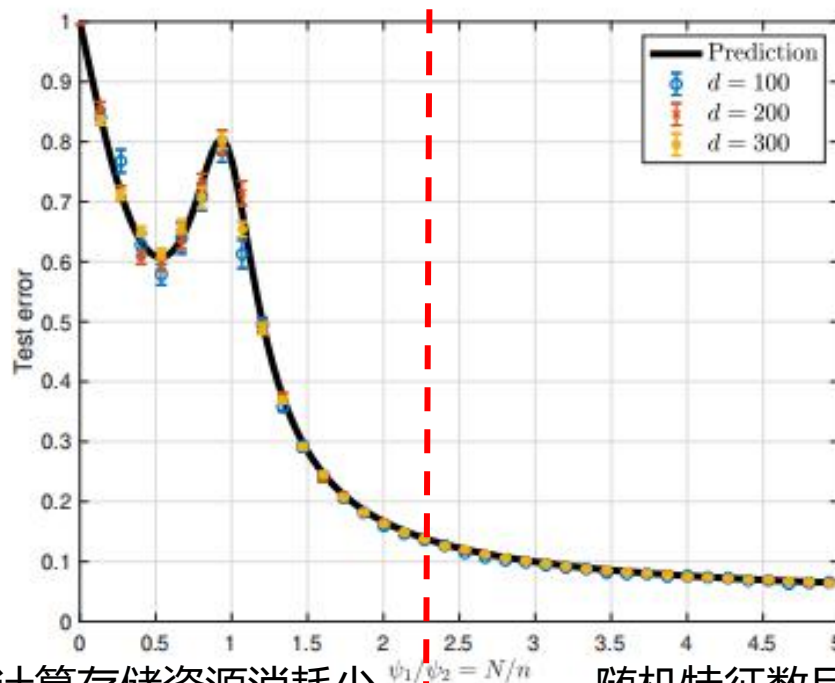
## 4 基于神经正切核低维假设的小样本学习研究

## 5 结论与展望

# 低比特量化随机特征的构建研究

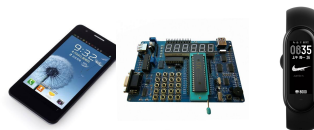


## 计算存储限制下的矛盾——为什么要研究低比特量化随机特征？



随机特征数目少，计算存储资源消耗少  
泛化性能变差

随机特征数目多，计算存储资源消耗多  
泛化性能变好



计算存储资源有限

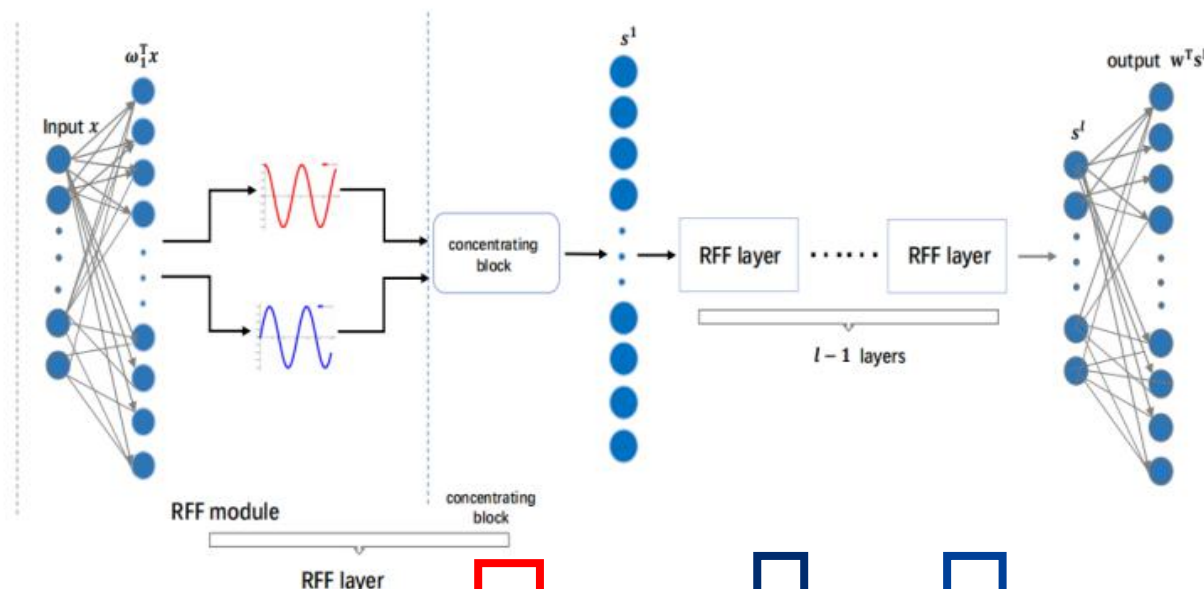
- [1] Mei S, Montanari A. The generalization error of random features regression: Precise asymptotics and the double descent curve CPAM 2019.  
[2] Avron H, Kapralov M, et al. Random Fourier features for kernel ridge regression: Approximation bounds and statistical guarantees[C]//ICML 2017

# 低比特量化随机特征的构建研究



## 计算存储限制下的矛盾——为什么要研究低比特量化随机特征？

DKR



$$\text{Memory Consumption} = 32 \times (2D_l^2 \times (l-1) + D_l \times D_S)$$

数据比特数目  
如果采用4比特、8比特？  
保证随机特征的数目

随机特征的数目



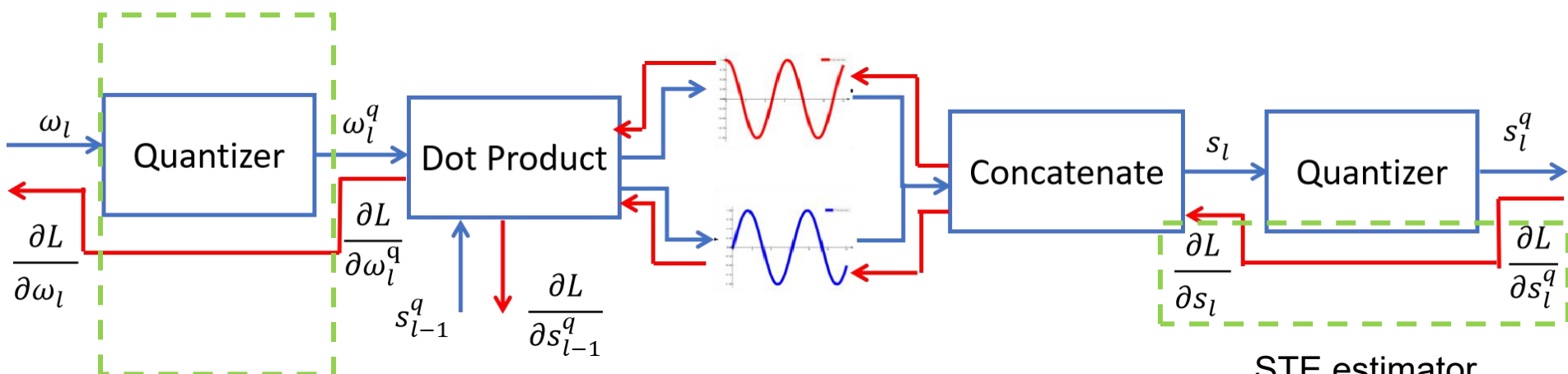
# 低比特量化随机特征的构建研究



## 低比特量化随机特征的构建过程

每层采样值矩阵 $\{\omega_i\}_{i=1}^l$ 和随机特征 $\{s_i\}_{i=1}^l$ 进行轻量化

Quantized RFF layer模块计算图

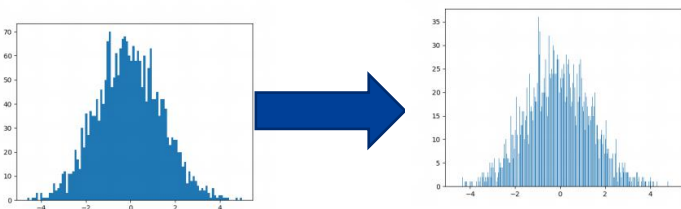


量化函数

STE estimator

$$\frac{\partial L}{\partial s_l} = \frac{\partial L}{\partial s_l^q}$$

解决量化函数求导过程中梯度弥散或者爆炸的现象



——→ 前向计算 (Feed Forward)

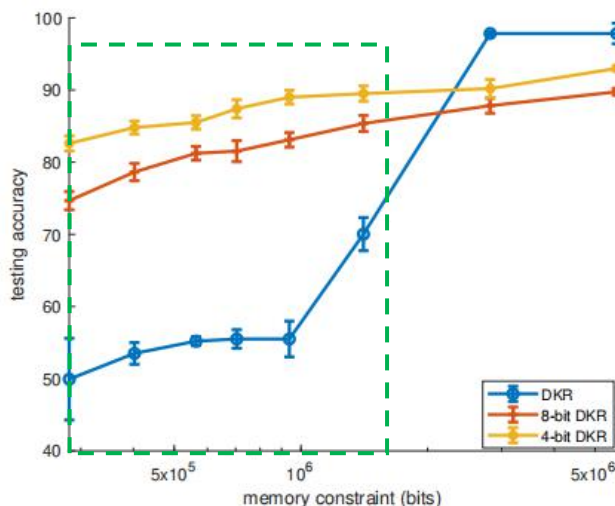
——→ 反向传播 (Back Propagation)



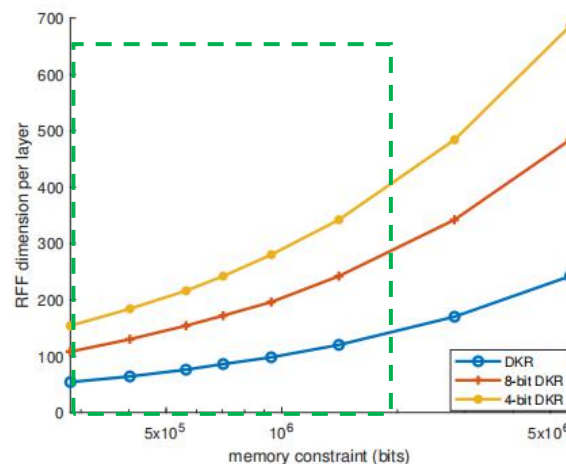
## 实验——解析低比特量化随机特征

$$2D_l = 2 \sqrt{\frac{\text{Memory Constraint}}{b \times 2 \times (l-1)}}$$

不同存储限制下采用32、4和8-bit DKR方法在EEG数据集上的性能



(a) EEG测试集分类准确率



(b) 每层允许最大的RFF维度

存储资源有限的情况下，量化比特数目↓，允许的RFF数目↑，核方法泛化性能↑

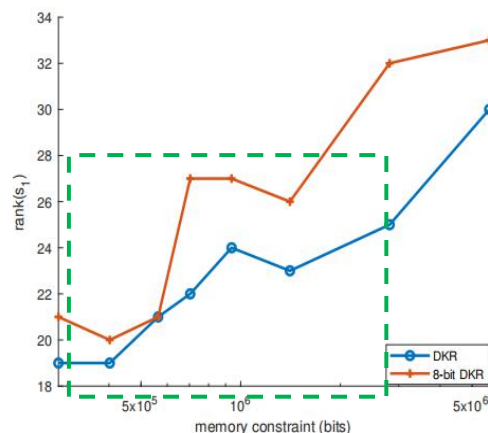
# 低比特量化随机特征的构建研究



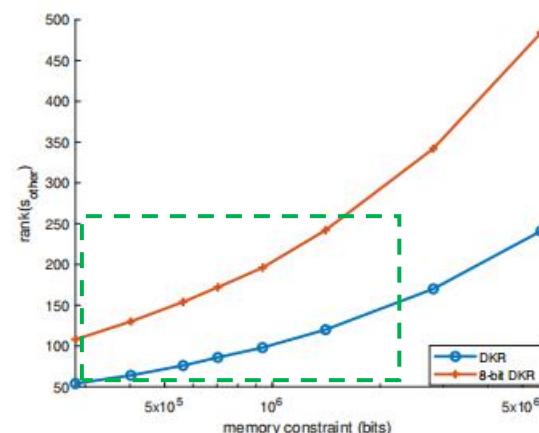
## 实验——解析低比特量化随机特征

随机特征的逼近  
矩阵的秩

第一层



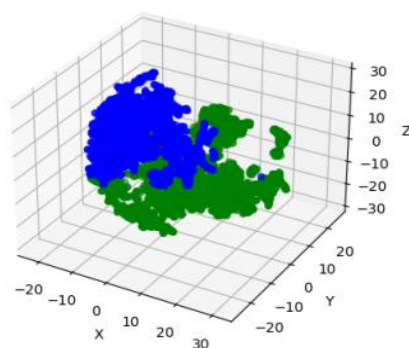
其余层



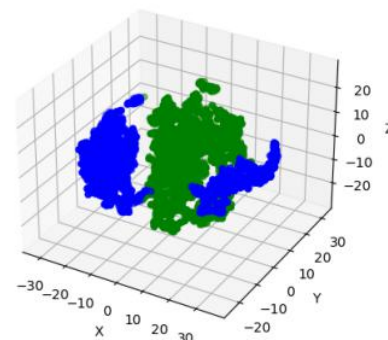
量化比特数目 ↓ →  $s$  ↑ →  $rank(Z)$  ↑ → 核方法的泛化能力 ↑

随机特征的可分性

32比特



8比特



量化比特数目 ↓ →  $s$  ↑ → 随机特征可分性 ↑ → 核方法的分类准确率 ↑



## 1 研究背景与意义

## 2 不定核随机特征构建方法研究

## 3 低比特量化随机特征的构建研究

## 4 基于神经正切核低维假设的小样本学习研究

## 5 结论与展望

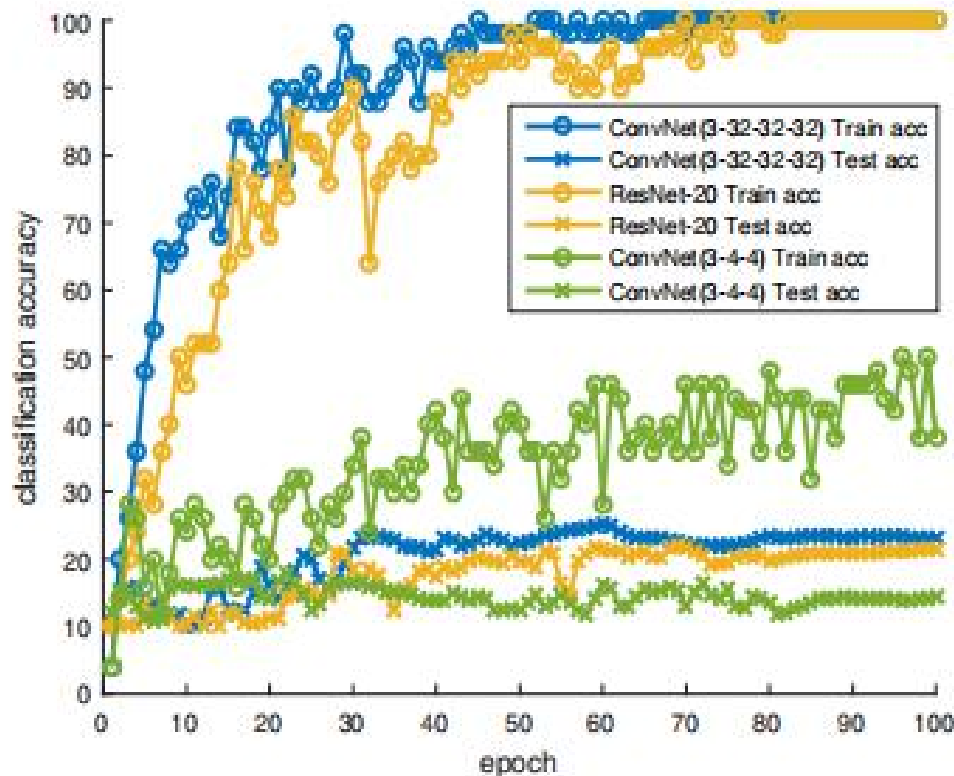
# 基于神经正切核低维假设的小样本学习研究



## 小样本学习问题建模

定义：训练样本数量极少（通常少于50）

带来的问题：严重的过拟合问题



训练集准确率为100%

测试集准确率为20%

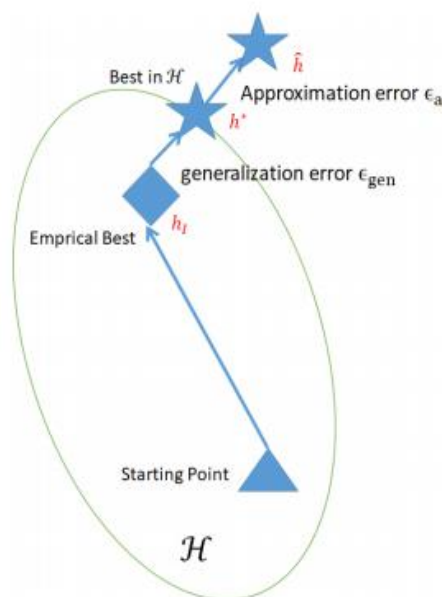


# 基于神经正切核低维假设的小样本学习研究

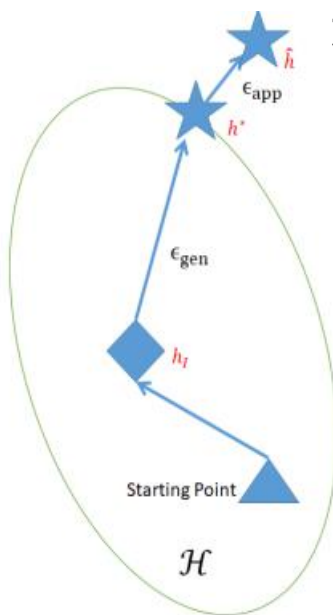
## 低维优化空间的构建——小样本学习问题的关键

$$\mathbb{E}[\mathcal{R}(h_I) - \mathcal{R}(\hat{h})] = \underbrace{\epsilon_{app}(\mathcal{H})}_{\text{逼近误差}} + \underbrace{\epsilon_{gen}(\mathcal{H}, \mathcal{X})}_{\text{泛化误差}}$$

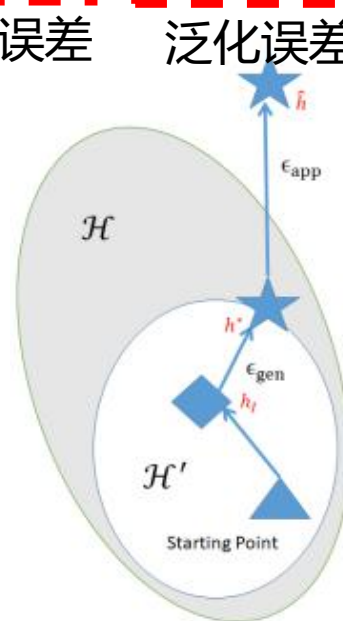
逼近误差      泛化误差



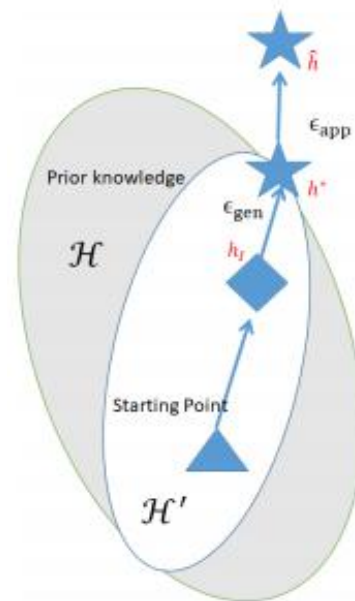
充足训练样本学习  
 $\epsilon_{app}$ 和 $\epsilon_{gen}$ 都小



小样本学习  
 $\epsilon_{gen}$ 增大, 过拟合



小样本学习中直接使用小  
模型或者模型剪枝  
 $\epsilon_{app}$ 增大



合适的低维优化空间?  
**基于NTK低维假设的低  
维优化空间**  
 $\epsilon_{app}$ 和 $\epsilon_{gen}$ 都小

根据**先验知识**构造合适的**低维优化空间**是解决小样本学习问题的关键。



## 基于NTK低维假设的优化空间构建——NTK相关理论

神经正切核 (NTK) —— 描述深度神经网络的**梯度下降**动态训练过程

$$\Theta = \Theta(\mathcal{X}, \mathcal{X}) = \nabla_{\theta} f(\theta, \mathcal{X}) \nabla_{\theta} f(\theta, \mathcal{X})^T$$

神经网络与核方法之间的桥梁

### 神经正切核与深度神经网络的联系

深度神经网络无限宽时

$$\nabla_t f(\theta, \mathcal{X}) = -\Theta(\mathcal{X}, \mathcal{X})(f(\theta, \mathcal{X}) - y)$$

达到稳态时，深度神经网络等价于核岭回归

$$f(\theta, x) \rightarrow \Theta(x, \mathcal{X})^T (\Theta(\mathcal{X}, \mathcal{X}))^{-1} y$$

[1] Jacot, Arthur, et al. "Neural tangent kernel: convergence and generalization in neural networks." NIPS (2018)

[2] Neal, Radford M. "Priors for infinite networks." Bayesian Learning for Neural Networks. 1996.

# 基于神经正切核低维假设的小样本学习研究

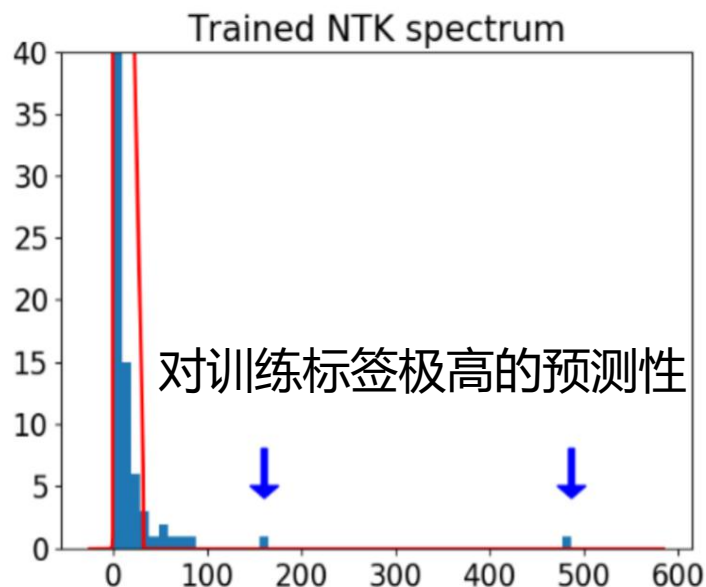


## 基于NTK低维假设的优化空间构建——NTK低维假设到优化子空间

NTK的低维逼近



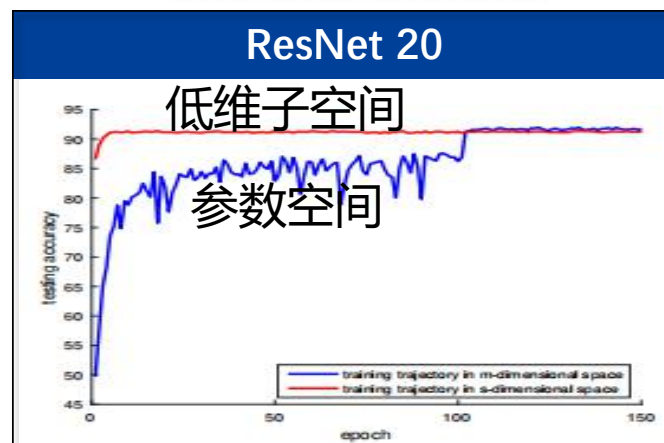
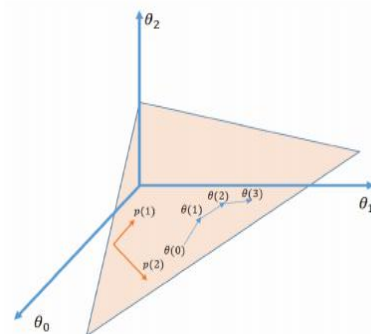
神经网络低维逼近特性



$$\dot{\theta}_t \approx - \underbrace{V_0 \tilde{V}_0}_{\text{未知变量}} \underbrace{(\tilde{\Sigma}_0 \tilde{U}_0^\top U_0^\top \nabla_{f^{\text{lin}}(X, \theta_t)} \mathcal{L})}_{\text{投影梯度}}$$

未知变量

投影梯度

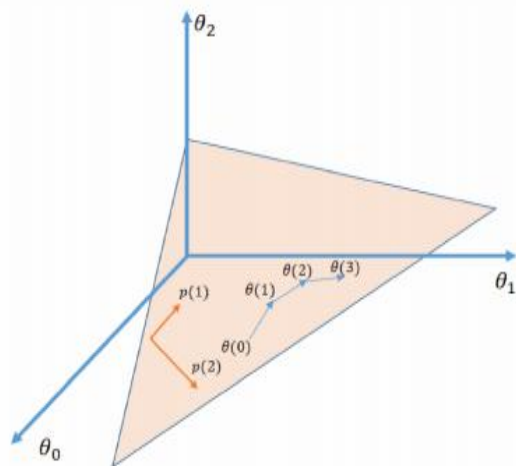


- [1] Li T, Tan L, et al. Low Dimensional Landscape Hypothesis is True: DNNs can be Trained in Tiny Subspaces[J]. arXiv preprint (submitted to TPAMI)
- [2] Fan, Zhou, and Zhichao Wang. "Spectra of the Conjugate Kernel and Neural Tangent Kernel for linear-width neural networks." NIPS 2020.

# 基于神经正切核低维假设的小样本学习研究



## 基于NTK低维假设的优化空间构建——不同训练样本容量下的性能

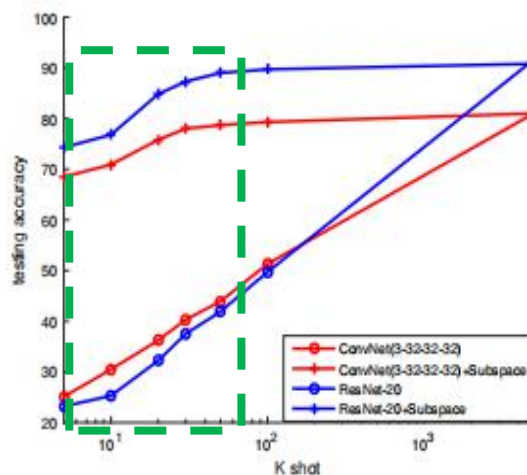


### P-SGD

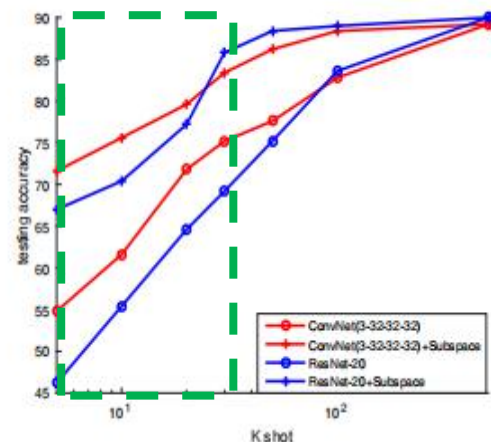
$$\theta(k+1) = \theta(k) - \eta P P^T \nabla_{\theta} \mathcal{L}$$

$P$ 根据原空间的优化轨迹进行PCA降维求取

CIFAR-10和CIFAR-100-small测试集分类准确率与训练样本数目关系



CIFAR-10



CIFAR-100-small

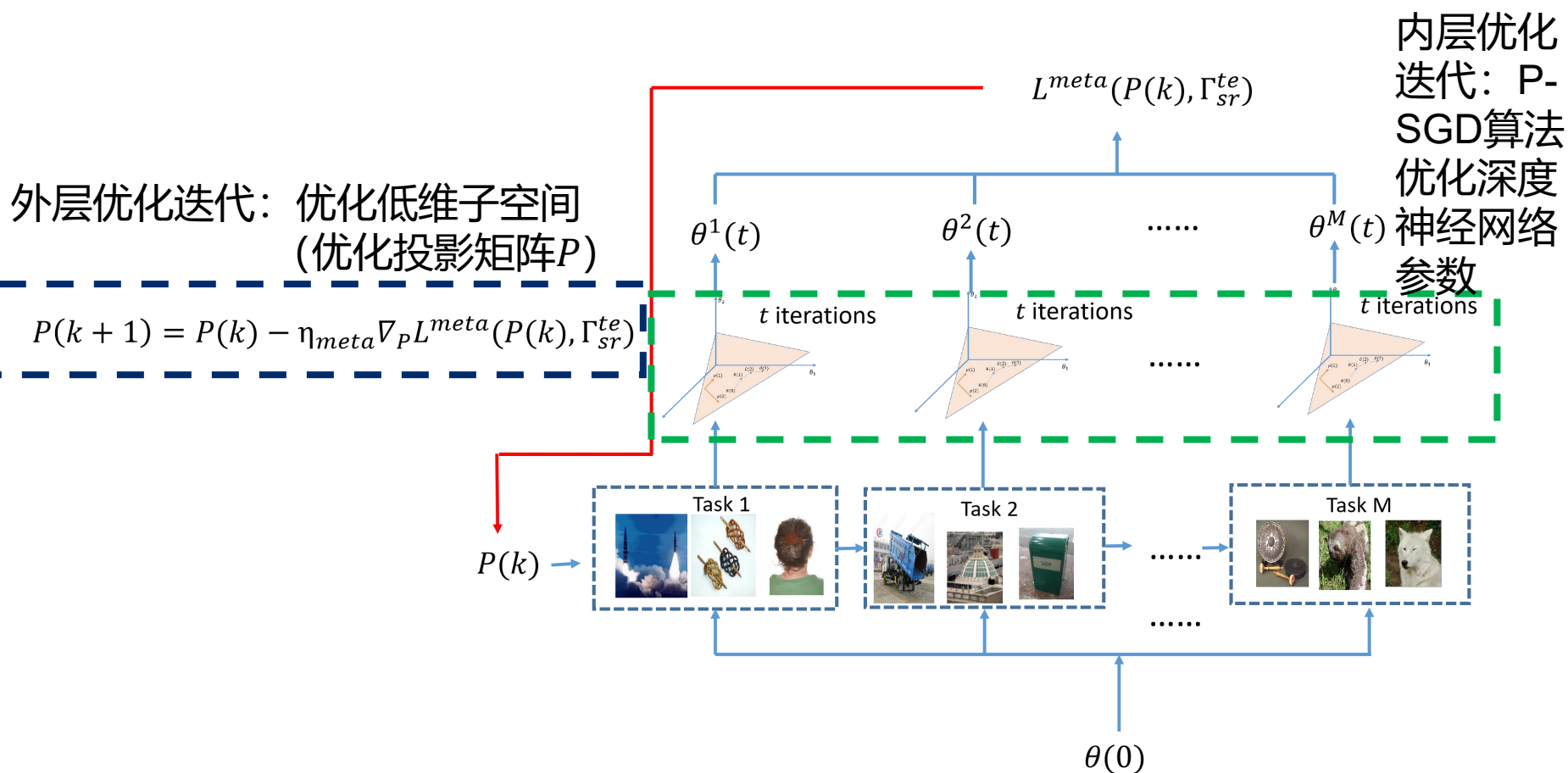
训练样本数量较少情况下,  
基于NTK的低维优化空间>原优化空间

# 基于神经正切核低维假设的小样本学习研究

## 基于NTK低维假设的优化空间构建——子空间的学习策略

问题：求取低维优化空间的先验知识过强（训练变化轨迹求取）

思路：学习低维优化空间，两层优化过程



# 基于神经正切核低维假设的小样本学习研究



## 实验——不同优化空间下在小样本分类问题上对比

| 数据集          | 优化空间                | 5-way 1-shot      | 5-way 5-shot      | 10-way 1-shot     | 10-way 5-shot     |
|--------------|---------------------|-------------------|-------------------|-------------------|-------------------|
| CIFAR-FS     | Original Space      | 37.95±7.79        | 53.40±7.29        | 25.75±3.67        | 39.52±6.76        |
|              | Pretrained Subspace | 40.66±6.00        | 56.25±4.90        | 21.12±3.17        | 26.96±4.09        |
|              | Meta Subspace       | <b>47.05±8.32</b> | <b>64.55±9.21</b> | <b>36.66±6.30</b> | <b>45.92±5.55</b> |
| MinilmageNet | Original Space      | 34.12±8.79        | 45.55±3.15        | 20.68±4.46        | 31.15±5.80        |
|              | Pretrained Subspace | 36.31±6.22        | 48.25±5.38        | 18.31±2.43        | 24.02±3.18        |
|              | Meta Subspace       | <b>44.07±7.65</b> | <b>56.84±8.73</b> | <b>30.06±5.42</b> | <b>37.92±4.75</b> |

Original Space——原参数空间    Pretrained Subspace——基于NTK低维假设和预训练得到的空间  
Meta Subspace——基于NTK低维假设和多任务学习调整得到的空间  
(训练集上的分类准确率达到100%)

### 结论:

Meta Subspace能够有效地缓解深度神经网络在小样本学习问题上出现的“过拟合”的情况，更加符合小样本场景设置。





## 1 研究背景与意义

## 2 不定核随机特征构建方法研究

## 3 低比特量化随机特征的构建研究

## 4 基于神经正切核低维假设的小样本学习研究

## 5 结论与展望

研究目标：解决随机特征构建存在的问题，拓宽随机特征的应用面

## 随机特征构建的角度

理论

不定核随机特征

定位：扩展随机特征的适用范围

展望：逼近方差进一步减小

低比特量化随机特征

定位：低计算存储资源落地场景下随机特征的构建

展望：应用于深层核的结构搜索

实践

基于NTK低维假设的小样本学习优化空间

定位：应用研究，解决深度神经网络在小样本学习上出现的“过拟合”的情况  
展望：即插即用，和其他小样本学习算法组合应用

# 攻读学位期间的学术成果及参与项目



## 学术论文:

**Luo Q**, Fang K, Yang J, Huang X. Towards Unbiased Random Features with Lower Variance For Stationary Indefinite Kernels[C]//2021 International Joint Conference on Neural Networks (IJCNN). IEEE, 2021: 1-8. **(EI, CCF-C 类, 交大 A 类会议, 已发表)**

Chu T, **Luo Q**, Yang J, Huang X. Mixed-precision quantized neural networks with progressively decreasing bitwidth[J]. Pattern Recognition, 2021, 111: 107647.

**(SCI 一区, CCF-B 类期刊, IF:7.74, 已发表)**

## 参与项目:

[1] 多层非正定核学习理论、算法及应用研究, 国家自然科学基金 (No. 61977046)

[2] 企业项目“电院-美敦力智慧医疗联合实验室”单目内窥镜图像深度重建

[3] 国防项目“面向人机协同的智能博弈方法研究”



# 欢迎评审老师批评指正！



上海交通大学  
SHANGHAI JIAO TONG UNIVERSITY