自适应深度学习异质性因子投资

Adaptive Neural Heterogenous Factor Investment

龚亦通

2025 年 12 月 30 日

- **动机:** 传统因子模型是线性且静态的，限制了其捕捉复杂市场动态的能力。
- **目标:** 开发基于神经网络的框架，以建模非线性因子交互和动态风险溢价。
- **方法论:**
  - 使用变分自编码器（VAE）学习因子的低维潜在流形。
  - 使用神经网络建模资产收益率的条件概率分布。
  - 资产间的定价机制同质，但基于因子的风险价格异质。
  - 使用混合高斯（MoG）以取代传统的噪声高斯假设。
  - 使用 Grad-CAM 进行因子归因分析。
  - 使用信息几何方法对因子状态和最优因子组合进行追踪和调仓。
- **预期结果:** 在市场状态转变期间实现更好的对冲和归因，并增强风险因子的可解释性，实现对 Barra 体系在非线性、非高斯环境下的推广。

目录

目录

什么是神经网络?

- 神经网络是一种受人脑启发的计算模型，由多层相互连接的节点（神经元）组成。
- 每个神经元处理输入数据并将输出传递到下一层。
- 神经网络通过训练可以学习数据中的复杂模式和关系。

数学表示为函数:

$$\mathbf{y} = f_\theta(\mathbf{x})$$

其中 $f_\theta : \mathbb{R}^m \mapsto \mathbb{R}^n$ 是神经网络函数，$\mathbf{x}$ 是输入，θ 表示网络的参数（权重和偏置）。

- 神经网络一大优势在于可以进行自动微分，从而高效计算复杂函数的导数。

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \frac{\partial f_\theta(\mathbf{x})}{\partial \mathbf{x}}$$

万能近似定理

定理：对于任何连续函数 $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ 和任意小的误差 $\epsilon > 0$，存在一个足够复杂的神经网络 f_θ，使得对于所有输入 $\mathbf{x}$ 在紧致集 $K \subset \mathbb{R}^m$ 上，有：

$$\|f(\mathbf{x}) - f_\theta(\mathbf{x})\| < \epsilon$$

这意味着神经网络可以以任意精度逼近任何连续函数。

神经网络（回归）vs 树模型（分类）

特征	神经网络	树模型
数学形式	$\mathbf{y} = f_\theta(\mathbf{x})$	递归分割输入空间
表达能力	能逼近任意连续函数	受限于分割方式
训练方法	反向传播与梯度下降	贪心算法
可解释性	较低	较高
适用场景	高维复杂数据	结构化数据
计算复杂度	$O(d^2N)$	$O(d^3N)$

表: 神经网络与树模型的比较

目录

因子特征工程（PCA）

在高维因子空间中，某些因子可能存在较强的相关性或冗余信息。为了简化模型并提高计算效率，我们可以通过因子合成的方法，将多个相关因子合成为一个新的综合因子。

最经典的方法则是主成分分析（PCA）。通过 PCA，我们可以找到数据中方差最大的方向，并将原始因子投影到这些方向上，从而得到一组新的无关因子。

具体地，假设我们有一个因子暴露矩阵 $B \in \mathbb{R}^{N\times K}$，则 PCA 的步骤如下：

- 计算因子暴露矩阵的协方差矩阵 $C = \frac{1}{N-1}B^\top B$。
- 计算协方差矩阵的特征值和特征向量。
- 选择前 m 个最大的特征值对应的特征向量，构成新的因子空间。
- 将原始因子暴露投影到新的因子空间，得到合成因子暴露矩阵 $B_{\text{new}} \in \mathbb{R}^{N\times m}$。

因子特征工程（PPCA）

概率主成分分析（Probabilistic PCA, PPCA）是一种基于概率模型的降维方法。与传统 PCA 不同，PPCA 通过引入潜在变量和高斯噪声模型，更加灵活地处理数据中的不确定性。具体地，PPCA 假设观测数据 $\mathbf{x} \in \mathbb{R}^D$ 由潜在变量 $\mathbf{z} \in \mathbb{R}^M$ 生成：

$$\mathbf{x} = W\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

其中 $W \in \mathbb{R}^{D\times M}$ 是投影矩阵，$\boldsymbol{\mu}$ 是均值向量，$\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 I)$ 是高斯噪声。通过最大化观测数据的似然函数，可以估计参数 W 和 σ^2，从而实现因子降维和合成。

变分自编码器（Variational Autoencoder, VAE）是一种基于神经网络的生成模型，能够学习数据的潜在表示。通过 VAE，我们可以将高维因子暴露映射到低维潜在空间，从而实现因子合成和降维。具体地，VAE 由编码器和解码器两部分组成：

- 编码器：将高维因子暴露 $\mathbf{x}$ 映射到潜在变量 $\mathbf{z}$ 的分布 $q_\phi(\mathbf{z}|\mathbf{x})$。
- 解码器：从潜在变量 $\mathbf{z}$ 生成高维因子暴露的重构分布 $p_\theta(\mathbf{x}|\mathbf{z})$。

通过最大化证据下界（ELBO）或最小化 KL 散度，我们可以同时优化编码器和解码器的参数，实现因子降维和合成。即：

$$\mathcal{L}(\theta, \phi; \mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

使用 VAE 方法可以用于伪数据生成，在数据较为稀少的情况下，生成模拟数据用于模型训练，从而增强因子模型的鲁棒性。

VAE 同样是 PPCA 方法的非线性非高斯的推广。

目录

线性因子（线性回归）模型假设

线性因子模型可以视作线性回归（Linear Regression）模型的扩展，假设资产收益可以通过一组共同因子来解释。这些因子代表了市场风险、行业风险或其他系统性风险来源。线性回归模型基于以下假设：

- 线性关系：因变量与自变量之间存在线性关系。
- 独立同分布：误差项独立且服从相同分布，通常假设为正态分布。
- 同方差性：误差项的方差是恒定的，不随自变量变化而变化。
- 无多重共线性：自变量之间不存在完全的线性相关关系。
- 同质性：数据点来自同一总体，具有相似的统计特性。

线性因子模型

传统的线性因子模型假设资产收益可以表示为因子收益的线性组合:

$$\mathbb{E}[R_{i,t}] = \alpha + \boldsymbol{\beta}_i^\top \boldsymbol{f}_t \implies R_{i,t} = \alpha + \boldsymbol{\beta}_i^\top \boldsymbol{f}_t + \epsilon_{i,t} = \begin{bmatrix} 1, \boldsymbol{\beta}_i^\top \end{bmatrix} \begin{bmatrix} \alpha \\ \boldsymbol{f}_t \end{bmatrix} + \epsilon_{i,t}$$

其中:

- $R_{i,t}$ 是资产 i 在时间 t 的收益率。
- α 是资产的截距项。
- $\boldsymbol{\beta}_i$ 是资产对因子的暴露度。
- $\boldsymbol{f}_t$ 是因子收益率向量。
- $\epsilon_{i,t}$ 是误差项，且可能为任意分布。

在 Barra 模型中，最优解可以表示为最大对数似然:

$$\alpha^*, \boldsymbol{f}_t^* = \arg\max_{\alpha, \boldsymbol{f}_t} \sum_{i=1}^{N} \log p(R_{i,t} | \boldsymbol{\beta}_i; \alpha, \boldsymbol{f}_t)$$

若误差项服从正态分布，则该最优解等价于最小二乘估计。

非线性因子模型

非线性因子模型使用非线性函数来捕捉资产收益与因子之间的复杂关系:

$$\mathbb{E}[R_{i,t}] = f_\theta(\boldsymbol{\beta}_i) \implies R_{i,t} = f_\theta(\boldsymbol{\beta}_i) + \epsilon_{i,t}$$

其中:

- f_θ 是一个非线性函数，能够捕捉因子之间的复杂交互作用。
- 其他符号与线性模型相同。

这种方法允许更灵活地建模资产收益的动态特性，适应非线性关系和因子之间的交互作用。其最优解同样可以通过最大化对数似然来获得:

$$\theta^* = \arg\max_\theta \sum_{i=1}^{N} \log p(R_{i,t}|\boldsymbol{\beta}_i;\theta)$$

使用神经网络拟合，可以加入正则化项或者更加灵活的惩罚项以防止过拟合。这是线性模型所不具备的优势。

非线性因子模型的因子收益率

在线性因子模型中，因子收益率通常通过线性回归估计。然而，在非线性因子模型中，资产 i 的因子收益率可以通过神经网络进行求导计算。

$$\mathbf{g}_{i,t}^{*} = \left.\frac{\partial \mathbb{E}[R_{i,t}]}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} = \left.\frac{\partial f_{\theta^*}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} \in \mathbb{R}^{D}$$

假设共有 N 个资产，则在时间 t 的因子收益率雅可比（Jacobian）矩阵为：

$$\frac{\partial \mathbf{r}_t}{\partial \boldsymbol{\beta}} = J_t^{*} = \begin{bmatrix} \mathbf{g}_{1,t}^{*} \\ \mathbf{g}_{2,t}^{*} \\ \vdots \\ \mathbf{g}_{N,t}^{*} \end{bmatrix} \in \mathbb{R}^{N\times D}$$

因子相互作用

非线性因子模型通过非线性函数 $f_\theta(\cdot)$ 捕捉因子对于超额收益率的共同作用。
雅可比向量描述了资产 i 的收益率对各个因子暴露的敏感度，即因子收益率。
而因子之间的相互作用可以通过雅可比矩阵的二阶导数（黑塞矩阵）来描述：

$$H^*_{i,t} = \left.\frac{\partial^2 \mathbb{E}[R_{i,t}]}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} = \left.\frac{\partial^2 f_{\theta^*}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} \in \mathbb{R}^{D \times D}$$

该矩阵的元素 $H^*_{i,t}[j,k]$ 表示因子 j 和因子 k 之间的交互作用对资产 i 收益率的影响。

将任意资产视作在因子空间中的点，其坐标由该资产对各个因子的暴露度决定。
使用二阶泰勒展开式，可以将任意资产的收益率拆解为以下部分：

$$r(\boldsymbol{\beta}) - r(\boldsymbol{\beta}_0) \approx \underbrace{\nabla_{\boldsymbol{\beta}} r(\boldsymbol{\beta}_0)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0)}_{\text{一阶项：线性因子收益}} + \underbrace{\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^\top H(\boldsymbol{\beta}_0)(\boldsymbol{\beta} - \boldsymbol{\beta}_0)}_{\text{二阶项：非线性因子收益}} + \epsilon$$

因为因子暴露本身已做过中心化处理，故可令 $\boldsymbol{\beta}_0 = 0$，则有：

$$r(\boldsymbol{\beta}) \approx r_0 + \underbrace{\mathbf{g}(0)^\top \boldsymbol{\beta}}_{\text{一阶项：线性因子收益}} + \underbrace{\frac{1}{2}\boldsymbol{\beta}^\top H(0)\boldsymbol{\beta}}_{\text{二阶项：非线性因子收益}} + \epsilon$$

定价机制同质性 vs 风险价格异质性

定价机制的同质性体现在：

- 所有资产共享相同的定价函数 $f_\theta(\cdot)$，即所有资产的收益都由同一非线性映射决定。
- 这种假设简化了模型结构，使得不同资产之间的比较和分析更加直接。

风险价格的异质性体现在：

- 所有资产共享同一定价函数，但由于其因子暴露位于不同状态点，定价函数对因子的局部敏感度（Jacobian）在资产之间存在差异。
- 这种异质性反映了资产对风险的不同敏感性，允许模型捕捉更复杂的市场动态。
- 通过这种方式，模型能够同时考虑市场的整体定价机制和个别资产的特定风险特征。

线性因子模型 vs 非线性因子模型总结

特征	线性因子模型	非线性因子模型
数学形式	$\mathbb{E}[R_{i,t}] = \alpha + \boldsymbol{\beta}_i^\top \boldsymbol{f}_t$	$\mathbb{E}[R_{i,t}] = f_\theta(\boldsymbol{\beta}_i)$
因子收益率计算	通过线性回归估计	通过神经网络求导计算
定价机制	同质性	同质性
风险价格	同质性	异质性
模型灵活性	受限于线性关系	能捕捉非线性关系和复杂交互作用
适用场景	适用于线性关系明显的市场	适用于复杂市场动态和非线性关系

表: 线性因子模型与非线性因子模型的比较

但是！

此时我们对于误差项一无所知，即 $\Sigma_\epsilon = Cov(\epsilon)$，实际建模时仍然存在较大困难。

但是！

此时我们对于误差项一无所知，即 $\Sigma_\epsilon = Cov(\epsilon)$，实际建模时仍然存在较大困难。

下一步：我们将尝试对资产收益率的概率分布直接建模，从而推断因子间的动态关系。

我们可以使用神经网络来建模资产收益率的概率密度函数 $p(r|\boldsymbol{\beta};\theta)$，从而捕捉更复杂的误差结构。具体地，我们可以定义一个神经网络 f_θ，其输出为资产收益率的对数概率密度：

$$\log p(r|\boldsymbol{\beta};\theta) = f_\theta(r,\boldsymbol{\beta})$$

通过最大化对数似然函数，我们可以估计参数 θ：

$$\theta^* = \arg\max_\theta \sum_{i=1}^{N} \log p(r_i|\boldsymbol{\beta}_i;\theta)$$

这种方法允许我们灵活地建模资产收益率的分布，捕捉非线性关系和复杂的误差结构。

我们可以使用神经网络来建模资产收益率的概率密度函数 $p(r|\boldsymbol{\beta};\theta)$，从而捕捉更复杂的误差结构。具体地，我们可以定义一个神经网络 f_θ，其输出为资产收益率的对数概率密度：

$$\log p(r|\boldsymbol{\beta};\theta) = f_\theta(r,\boldsymbol{\beta})$$

通过最大化对数似然函数，我们可以估计参数 θ：

$$\theta^* = \arg\max_\theta \sum_{i=1}^{N} \log p(r_i|\boldsymbol{\beta}_i;\theta)$$

这种方法允许我们灵活地建模资产收益率的分布，捕捉非线性关系和复杂的误差结构。

但是，暴力设置神经网络还是不行，无法约束概率密度函数积分为 1。
此时，需要引入混合高斯模型（Mixture of Gaussians, MoG）来进行改造。

混合高斯模型

混合高斯模型（Mixture of Gaussians, MoG）是一种概率模型，用于表示数据点由多个高斯分布生成的情况。其概率密度函数表示为:

$$p(r|\boldsymbol{\beta};\theta)=\sum_{k=1}^{K}\pi_k(\boldsymbol{\beta};\theta)\mathcal{N}(r|\mu_k(\boldsymbol{\beta};\theta),\sigma_k^2(\boldsymbol{\beta};\theta))$$

其中:

- K 是高斯分布的数量。
- $\pi_k(\boldsymbol{\beta};\theta)$ 是第 k 个高斯分布的混合权重，满足 $\sum_{k=1}^{K}\pi_k(\boldsymbol{\beta};\theta)=1$。
- $\mu_k(\boldsymbol{\beta};\theta)$ 和 $\sigma_k^2(\boldsymbol{\beta};\theta)$ 分别是第 k 个高斯分布的均值和方差。

混合高斯分布在选取合适的 K 情况下，可以逼近任意复杂的连续概率分布。

此时，我们就可以训练 $f_\theta:\mathbb{R}^D\mapsto\mathbb{R}^{3K}$ 这样的一个神经网络，用于输出混合高斯模型的参数。

最大化对数似然函数以估计参数 θ:

$$\theta^*=\arg\max_\theta\sum_{i=1}^{N}\log p(r_i|\boldsymbol{\beta}_i;\theta)=\arg\max_\theta\sum_{i=1}^{N}\log\left(\sum_{k=1}^{K}\pi_k(\boldsymbol{\beta}_i;\theta)\mathcal{N}(r_i|\mu_k(\boldsymbol{\beta}_i;\theta),\sigma_k^2(\boldsymbol{\beta}_i;\theta))\right)$$

资产收益率均值与方差

在混合高斯模型中，资产 i 的收益率均值和方差可以通过以下公式计算：

$$\mathbb{E}[R_{i,t}] = \sum_{k=1}^{K} \pi_k(\boldsymbol{\beta}_{i,t};\theta^*)\mu_k(\boldsymbol{\beta}_{i,t};\theta^*)$$

$$Var(R_{i,t}) = \sum_{k=1}^{K} \pi_k(\boldsymbol{\beta}_{i,t};\theta^*)\left(\sigma_k^2(\boldsymbol{\beta}_{i,t};\theta^*) + \mu_k^2(\boldsymbol{\beta}_{i,t};\theta^*)\right) - (\mathbb{E}[R_{i,t}])^2$$

在混合高斯模型中，资产 i 的因子收益率可以通过对混合高斯分布的均值进行求导计算：

$$\mathbf{g}^*_{i,t} = \frac{\partial \mathbb{E}[R_{i,t}]}{\partial \boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} = \frac{\partial}{\partial \boldsymbol{\beta}} \left(\sum_{k=1}^{K} \pi_k(\boldsymbol{\beta}; \theta^*) \mu_k(\boldsymbol{\beta}; \theta^*) \right) \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} \in \mathbb{R}^D$$

假设共有 N 个资产，则在时间 t 的因子收益率雅可比（Jacobian）矩阵为：

$$\frac{\partial \mathbf{r}_t}{\partial \boldsymbol{\beta}} = J^*_t = \begin{bmatrix} \mathbf{g}^*_{1,t} \\ \mathbf{g}^*_{2,t} \\ \vdots \\ \mathbf{g}^*_{N,t} \end{bmatrix} \in \mathbb{R}^{N \times D}$$

对于单一资产的二阶导数（黑塞矩阵）同样可以进行类似计算，以实现对于单一资产收益率的因子间的相互作用分析。

$$H^*_{i,t} = \frac{\partial^2 \mathbb{E}[R_{i,t}]}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \left(\sum_{k=1}^{K} \pi_k(\boldsymbol{\beta}; \theta^*) \mu_k(\boldsymbol{\beta}; \theta^*) \right) \bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_{i,t}} \in \mathbb{R}^{D \times D}$$

同理，可以对单一资产收益率不确定性进行类似分析。

目录

投资组合

假设我们有一个由 N 个资产组成的投资组合，其权重向量为 $\mathbf{w} = [w_1, w_2, \ldots, w_N]^\top$。则该投资组合的收益率 r_p 可以表示为：

$$r_p = \sum_{i=1}^{N} w_i R_i = \mathbf{w}^\top \mathbf{r}$$

其中 R_i 是资产 i 的收益率。
投资组合的预期收益率和方差分别为：

$$\mathbb{E}[r_p] = \mathbf{w}^\top \mathbb{E}[\mathbf{r}]$$

$$Var(r_p) = \mathbf{w}^\top Cov(\mathbf{r}) \mathbf{w}$$

因为已知各个资产的分布，可以直接推算期望值与方差。计算协方差工程难度较大，这里就不做展开。

此时投资组合预期收益率的因子敏感度（雅可比）可以表示为：

$$
\begin{aligned}
\boldsymbol{g}_j^{\text{port}} &= \frac{\partial \mathbb{E}[r_p]}{\partial \boldsymbol{\beta}_j^{\text{port}}} \\
&= \sum_{i=1}^{N} \left(\frac{\partial \mathbb{E}[r_p]}{\partial \mathbb{E}[r_i]} \cdot \frac{\partial \mathbb{E}[r_i]}{\partial \beta_{ij}} \cdot \frac{\partial \beta_{ij}}{\partial \beta_j^{\text{port}}} \right) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{N} \left(\frac{\partial \pi_{ik}}{\partial \beta_{ij}} \mu_{ik} + \pi_{ik} \frac{\partial \mu_{ik}}{\partial \beta_{ij}} \right)
\end{aligned}
$$

在此基础上，我们可以进一步计算投资组合收益率的二阶导数（黑塞矩阵）：

$$
H_{j,l}^{\text{port}} = \frac{\partial^2 \mathbb{E}[r_p]}{\partial \boldsymbol{\beta}_j^{\text{port}} \partial \boldsymbol{\beta}_l^{\text{port}}}
$$

该矩阵的元素 $H_{j,l}^{\text{port}}$ 表示因子 j 和因子 l 之间的交互作用对投资组合收益率的影响。

同理，可以对投资组合收益率的方差进行类似的因子敏感度和交互作用分析。使用链式法则可以对组合夏普率等指标进行类似的归因分析。

Grad-CAM 因子归因/挖掘

Grad-CAM（Gradient-weighted Class Activation Mapping）是一种用于解释神经网络决策的技术。它通过计算输出对输入特征的梯度，生成热力图，突出显示对模型预测贡献最大的区域。

在因子归因中，我们可以使用 Grad-CAM 来识别哪些因子对资产收益率的预测贡献最大。具体步骤如下：

- 计算组合收益率/组合标准差预测对各因子暴露的梯度。
- 使用这些梯度生成热力图，显示每个因子对预测的贡献。
- 分析热力图以识别关键因子，帮助投资决策。

算法解释：https://zhuanlan.zhihu.com/p/479485138

相较于使用泰勒公式进行因子归因，Grad-CAM 方法能够更直观地展示因子间非线性耦合的重要性，并且不依赖于各阶导数的计算，适用于复杂的非线性模型。

同理，可以对资产权重进行类似的归因分析。

该方法同时可以实现非线性耦合因子的挖掘。

目录

信息几何简介

信息几何（Information Geometry）是一种将黎曼几何方法应用于概率论和统计学的学科。它通过将概率分布视为流形上的点，利用几何工具来分析和理解统计模型的性质。主要概念包括：

- **概率流形（Statistical Manifold）**：将参数化的概率分布视为流形上的点。
- **费舍信息度量（Fisher Information Metric）**：用于测量流形上点之间的距离，定义为：

$$g_{ij}(\beta) = \mathbb{E}\left[\frac{\partial \log p(X|\beta)}{\partial \beta_i}\frac{\partial \log p(X|\beta)}{\partial \beta_j}\right]$$

- **测地线（Geodesics）**：流形上连接两点的最短路径，表示参数变化的最优路径。

信息几何在机器学习、统计推断和优化等领域有广泛应用，帮助理解复杂模型的结构和行为。

因子收益率的流形表示

在信息几何框架下，每组因子暴露对应着一个收益率分布 $p_\theta(r|\beta)$，则我们可以构造流形

$$\mathcal{M} = \{p_\theta(r|\beta) : \beta \in \mathbb{R}^D\},$$

其中每个点表示特定因子暴露下的收益率分布。

此时信息几何中的费舍信息度量可以用来衡量不同因子暴露下收益率分布之间的差异。
具体地，费舍信息矩阵定义为：

$$g_{ij}(\beta) = \mathbb{E}\left[\frac{\partial \log p_\theta(r|\beta)}{\partial \beta_i} \frac{\partial \log p_\theta(r|\beta)}{\partial \beta_j}\right]$$

该矩阵度量了因子暴露空间中微小变化对收益率分布的影响。
此时，期望收益率函数为：

$$f(\beta) = \mathbb{E}[R|\beta] = \int r\, p_\theta(r|\beta)\, dr$$

此时流形的曲率可以反映因子暴露空间中收益率分布的复杂性和非线性关系。数学上，曲率张量 R_{ijkl} 可以通过费舍信息矩阵的二阶导数计算得到。

因子套利机会

在欧式几何中，套利机会可以通过因子暴露空间中的环积分来确定。然而，在信息几何框架下，我们需要考虑概率分布流形上的路径积分。

具体地，考虑因子暴露空间中的闭合路径 C，则在信息几何中，套利机会可以表示为：

$$\oint_C df(\beta)$$

其中 $df(\beta)$ 是期望收益率函数 $f(\beta)$ 的外微分。

如果该路径积分不为零，则表示存在套利机会，因为沿着闭合路径，投资者可以实现无风险收益。

另外，该套利机会同样体现在流形曲率的非平坦性上。若流形具有非零曲率，则说明因子暴露空间中存在复杂的几何结构，可能导致套利机会的出现。

详见：Geometric Arbitrage Theory and Market Dynamics。

这种方法利用信息几何的工具，提供了一种新的视角来分析因子暴露空间中的套利机会。

因子拥挤度

因子拥挤度（Factor Crowding）是指大量投资者集中持有某些因子暴露，导致这些因子的风险溢价下降或收益率波动加剧的现象。在信息几何框架下，我们可以通过分析因子暴露空间中的概率分布流形来量化因子拥挤度。

具体地，因子拥挤度可以通过计算流形上矢量场的散度来衡量。散度反映了流形上因子暴露的集中程度，较高的散度值表示因子暴露较为分散，而较低的散度值则表明因子暴露较为集中。具体地，给定一个矢量场 V，其散度定义为：

$$\mathrm{div}(V) = \frac{1}{\sqrt{|g|}} \partial_i (\sqrt{|g|} V^i)$$

投资组合优化

在信息几何框架下，我们不再使用马科维茨均值-方差优化方法，而是通过在概率分布流形上寻找最优路径来实现投资组合优化。

首先，投资组合优化问题可以表示为在流形 $\mathcal{M}$ 上寻找一组因子暴露 β^*，使得预期收益率最大化，同时满足风险约束：

$$\beta^* = \arg\max_{\beta} f(\beta) \quad \text{subject to} \quad Var(r_p|\beta) \leq \sigma^2_{\max}$$

在流形中，我们可以风险约束表示为测地线距离的限制，从而确保投资组合在风险可控的范围内。

$$d_G(\beta, \beta_{\text{benchmark}}) \leq d_{\max}$$

基于测地线方向的动态优化

在黎曼流形中，测地线距离 $d_G(\beta_1, \beta_2)$ 表示流形上两点之间的最短路径长度。我们可以利用测地线距离来动态调整投资组合的因子暴露，以适应市场变化。

欧式空间中，梯度可以给出最快上升方向；而在黎曼流形中，测地线方向则表示最快上升路径。因此，我们可以通过沿测地线方向调整因子暴露，实现投资组合的动态优化。即为：

$$\mathrm{grad}_{\mathcal{M}} f(\beta) = G^{-1}(\beta)\nabla f(\beta)$$

在具体实施上，可以参考各类优化器在黎曼流形上的推广方法，如 Riemannian Gradient Descent 等，实现对因子暴露的动态调整。

$$\begin{aligned} \beta_{t+1} &= \beta_t + \eta_t \nabla f(\beta_t) \quad (\text{欧式}) \\ \beta_{t+1} &= R_{\beta_t}\left(\eta_t \, \mathrm{grad}_{\mathcal{M}} f(\beta_t)\right) \quad (\text{黎曼}) \end{aligned}$$

如果引入一些自适应的学习率调整方法，如 AdaGrad、Adam、RMSprop，可以进一步提升优化效果。

基于风险偏好的动态优化

我们可以实时监测市场的风险偏好变化，并根据这些变化动态调整投资组合的优化目标。例如市场对于部分因子的风险偏好上升，流形在这些因子方向上曲率发生变化，此时我们可以调整优化目标函数 $f(\beta)$，以反映新的风险偏好。例如，可以引入一个风险偏好权重 λ，调整预期收益率函数为：

$$f_\lambda(\beta) = \mathbb{E}[R|\beta] - \lambda \cdot Var(R|\beta)$$

通过动态调整 λ，我们可以使投资组合更好地适应市场的风险偏好变化，从而实现更有效的风险管理和收益优化。

总结

- 因子合成方法（PCA、PPCA、VAE）有效地降低了因子空间的维度，简化了模型结构，提高了计算效率，同时保留了关键的风险信息。
- 非线性因子模型通过神经网络和混合高斯模型，能够更灵活地捕捉资产收益率的复杂分布和因子间的非线性关系。
- 利用信息几何框架，我们可以将因子暴露空间视为概率分布流形，利用几何工具分析因子动态变化、套利机会和因子拥挤度。
- 基于测地线方向和风险偏好的动态优化方法，为投资组合管理提供了新的思路，能够更好地适应市场变化，实现风险控制和收益最大化。

交流时间