Задача (Байесовский метод главных компонент).

Рассмотрим вероятностную модель метода главных компонент, считая, что для ~~каж~~ каждого $x_i \in \mathbb{R}^n$ $\exists$ описание $z_i \in \mathbb{R}^d$ в признаковом пространстве меньшей размерности, причем $x_i = Wz_i + \mu + \varepsilon_i$, где $\mu \in \mathbb{R}^n$ — смещение, $\varepsilon_i \in \mathbb{R}^n$ — шумовой вектор.

] Есть выборка $X = [x_1, \dots, x_m]$ независимых объектов.

Пусть $p(z_i) = \mathcal{N}(z_i|0, I)$, $p(\varepsilon_i) = \mathcal{N}(\varepsilon_i|0, \sigma^2 I)$

считая W, σ^2, μ — неизвестными параметрами задачи, а d — фиксирован.

т.к. Z не зависит от W, μ, σ

а) Вычислить $p(X, Z|W, \mu, \sigma)$

$p(X, Z|W, \mu, \sigma) = \cancel{\ldots} = p(X|W, \mu, \sigma, Z)\cdot p(Z) =$

выборка независима и $x_i - Wz_i - \mu = \varepsilon_i \in \mathcal{N}(0, \sigma^2 I)$

$= \prod_{i=1}^{m} \mathcal{N}(x_i|\mu + Wz_i, \sigma^2 I)\, \mathcal{N}(z_i, 0, I)$

б) Найти $p(X|W, \mu, \sigma)$ $\mathbb{E}[x] = \mathbb{E}[Wz + \mu + \varepsilon] = \mu$; $\operatorname{cov}[x] = \mathbb{E}[(Wz+\varepsilon)(Wz+\varepsilon)^T] = \mathbb{E}[Wzz^T W] + \mathbb{E}[\varepsilon\varepsilon^T] = WW^T + \sigma^2 I$

$p(X|W, \mu, \sigma) = \prod_{i=1}^{m} p(x_i|W, \mu, \sigma) = \prod_{i=1}^{m} \mathcal{N}(x_i|\mu, WW^T + \sigma^2 I)$

в) С помощью EM-алгоритма решить задачу нахождения наиболее правдоподобных оценок W, μ, σ, т.е. $p(X|W, \mu, \sigma) \to \max_{W, \mu, \sigma}$

Получив итеративные формулы пересчета для E и M шагов.

Каково апостериорное распределение $p(z_i|x_i, W, \sigma, \mu)$.

Как изменить вероятностную модель, чтобы учесть, что в данных есть пропуски?

$F(q, W, \mu, \sigma) = \log p(X|W, \mu, \sigma) - \sum_{i=1}^{m} D_{KL}(q(z_i)\,\|\,p(z_i|x_i, W, \mu, \sigma))$

Решаем задачу правдоподобности, это следует из:

$\max_{W,\mu,\sigma} p(X|W, \mu, \sigma) = \max_{W,\mu,\sigma} \log(p(X|W, \mu, \sigma))$ и того, что $\log p(X|W, \mu, \sigma) \geq \sum_{i=1}^{m} [\mathbb{E}_{q(z_i)} \log p(x_i|W, \mu, \sigma) - D_{KL}(q(z_i)\|p(z_i|x_i, W, \mu, \sigma))] =$

$= \log p(X|W, \mu, \sigma) - \sum_{i=1}^{m} D_{KL}(q(z_i)\|p(z_i|x_i, W, \mu, \sigma))$

$\|$

$F(q, W, \mu, \sigma)$

E-шаг: $q(z_i) = p(z_i|x_i, W, \mu, \sigma)$

Апостериор на z_i — нормальное, т.к. само правдоподобие нормальное. Тогда апостериор будет иметь следующие параметры (мы их уже считали в прошлой домашке).

$$q(z_i) = \mathcal{N}(z_i \mid m_i, C) \qquad C = \left(\frac{1}{\sigma^2} W^T W + I\right)^{-1}$$

$$m_i = \frac{1}{\sigma^2} C W^T (x_i - \mu)$$

M-шаг: $$\mathbb{E}_{q(z)} \log(p(X, Z \mid W, \mu, \sigma)) = \Phi(W, \mu, \sigma) = \sum_{i=1}^{m} \mathbb{E}_{q(z)}\left[-\frac{1}{2} z_i^T z_i - \frac{1}{2\sigma^2}(Wz_i + \mu - x_i)^T \cdot (Wz_i + \mu - x_i) - \frac{n}{2}\log\sigma^2\right] - \frac{m(d+n)}{2}\log(2\pi)$$

$$\frac{\partial \Phi}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{m} \mathbb{E}_{q(z)}[W z_i + \mu - x_i] = 0 \Rightarrow \mu = \frac{1}{m} \sum_{i=1}^{m} x_i - W \mathbb{E}_{q(z_i)} z_i \Rightarrow$$

$$\Rightarrow \mu^* = \sum_{i=1}^{m} \frac{x_i}{m}$$

$$\frac{\partial \Phi}{\partial W} = \sum_{i=1}^{m} \mathbb{E}_{q(z)}[(W z_i + \mu - x_i) z_i^T] = \sum_{i=1}^{m} [W \mathbb{E}_{q(z_i)} z_i z_i^T + (\mu - x_i) \mathbb{E} z_i^T] \Rightarrow$$

$$W^* = \left(\sum_{i=1}^{m} (x_i - \mu^*) \mathbb{E}_{q(z_i)} z_i^T\right) \cdot \left(\sum_{i=1}^{m} \mathbb{E} z_i z_i^T\right)^{-1}$$

$$\frac{\partial \Phi}{\partial \sigma} = \frac{1}{2\sigma^4} \sum_{i=1}^{m} \mathbb{E}_{q(z)}[(W z_i + \mu - x_i)^T (W z_i + \mu - x_i)] - \frac{n \cdot m}{2\sigma^2} = 0 \Rightarrow$$

$$\sigma_*^2 = \frac{1}{n \cdot m} \sum_{i=1}^{m} \left(\mathbb{E}[z_i^T W_*^T \cdot W_* z_i] + (\mu_* - x_i)^T (\mu_* - x_i) + 2\mathbb{E}[z_i^T W_*^T (\mu_* - x_i)]\right)$$

$$\mathbb{E} z_i = m_i = \frac{1}{\sigma^2} C W^T (x_i - \mu) \qquad \mathbb{E} z_i z_i^T = C + m_i \cdot m_i^T$$

$$\mathbb{E}\, z_i^T \cdot W_*^T \cdot W_* \cdot z_i = \mathrm{tr}(W_*^T W_* C) + m_i^T W_*^T W_* m_i$$

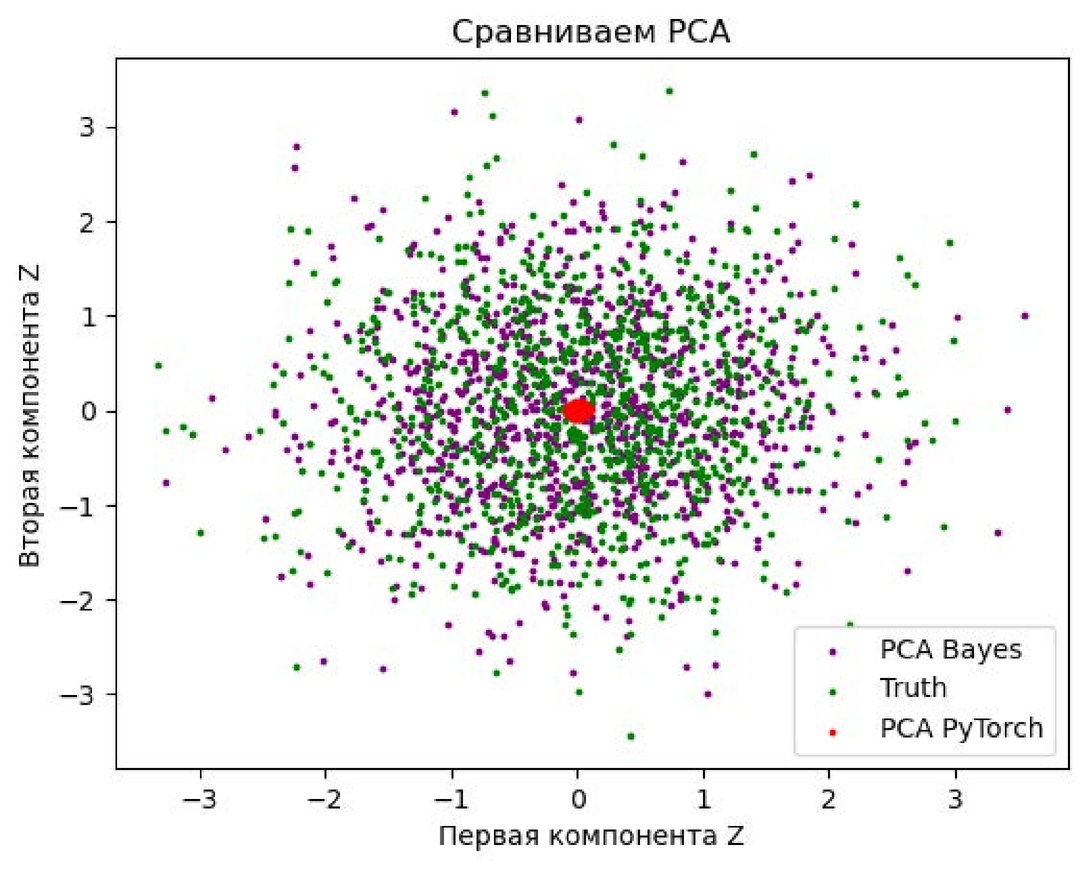
Если в данных имеются пропуски, то разбиваем вектор на два подвектора и по каждому отдельно предсказываются вероятности т.к. очевидно, что они будут у них разные (это было видно на практике, только там были outliers/выбросы).

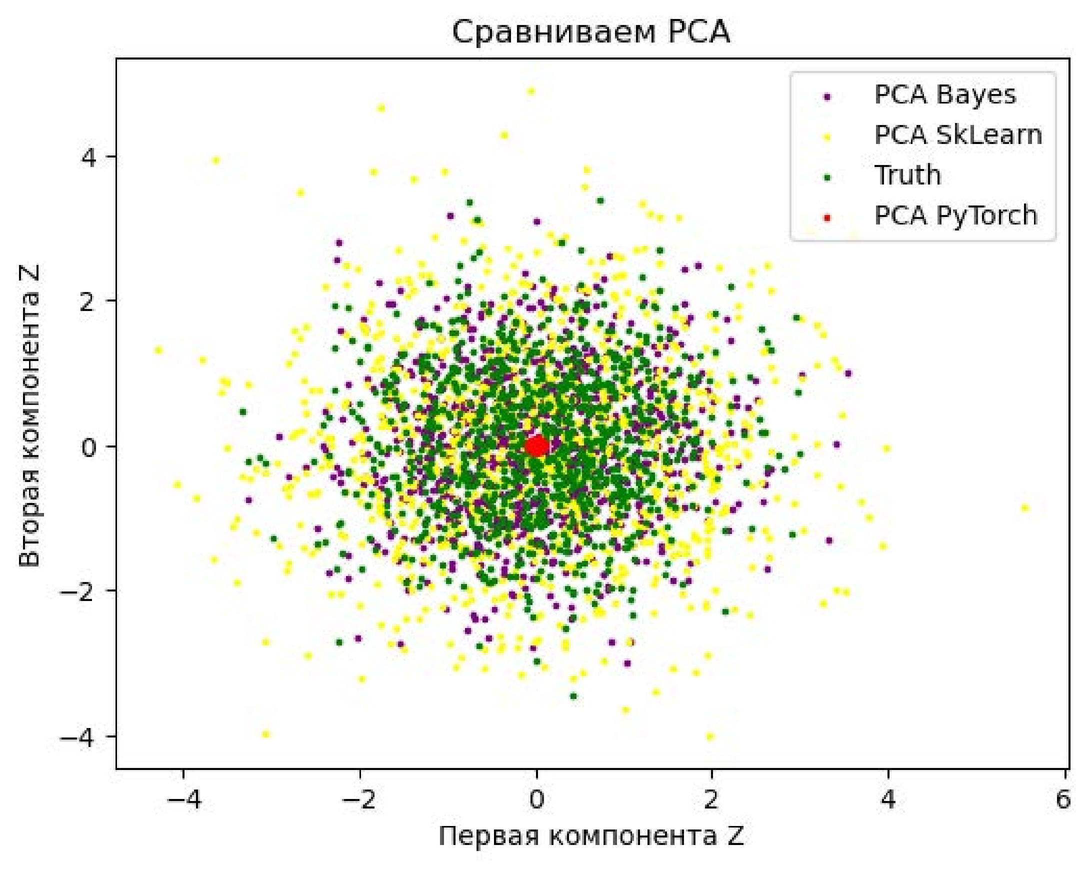
$$p(X_{full}, X_{skip} \mid W, \mu, \sigma) = \prod_{i=1}^{m} \mathcal{N}(x^i_{full} \mid W_{full} z_i + \mu_{full}, \sigma^2 I) \cdot \mathcal{N}(x^i_{skip} \mid W_{skip} z_i + \mu_{skip}, \sigma^2 I) \cdot \mathcal{N}(z_i \mid 0, I)$$

Теперь у нас есть $W_{full}, \mu_{full}, W_{skip}, \mu_{skip}$

Для новых параметров точно можно найти $q(z_i, x^i_{skip})$ – аналогично $q(z_i, x^i)$, и точно выписать формулы.

Баесовский PCA превосходит и реализацию обычного PCA из sklearn, и из PyTorch. По сравнению с обеими реализациями из коробки, он лучше приближает исходное распределение. PyTorch-овский вариант же автоматически выдает 6 компонент, а не 2, а в случае выбора двух компонент он выдает что-то концентрированное около нуля, что далеко от исходного распределения.


Сравниваем PCA
Вторая компонента Z
Первая компонента Z
PCA Bayes
Truth
PCA PyTorch


Сравниваем PCA
PCA Bayes
PCA SkLearn
Truth
PCA PyTorch
Вторая компонента Z
Первая компонента Z
4
2
0
−2
−4
−4
−2
0
2
4
6

д) (Автоматическое определение числа компонент).

$\alpha = n$ и введём априорное распределение на $W = [w_1, \dots, w_n]$

$$p(W|\alpha) = \prod_{i=1}^{n} \left(\frac{\alpha_i}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\alpha_i}{2} w_i^T w_i\right)$$

Если $\alpha_i \to \infty$, то $w_i^T w_i \to 0$, w_i — столбец W, т.е. исключаем соответств. компоненту из разложения $x_i = W z_i + \varepsilon + \mu$, т.е. как в PCA

С помощью вариац. EM-алгоритма решить задачу $p(X|\mu, \sigma, \alpha) \to \max\limits_{\alpha, \mu, \sigma}$

Правдоподобие: $p(X, Z, W|\mu, \alpha, \sigma) = p(X|Z, W, \mu, \alpha, \sigma)\, p(W|\alpha) \cdot p(Z)$

E-шаг: $q(W, z_1, \dots, z_m) = \prod_{i=1}^{m} q(z_i) \cdot q(W)$

$$\log q(z_i) = \mathbb{E}_{q(\neq z_i)} \log p(X, Z, W|\mu, \alpha, \sigma) \propto -\frac{1}{2} \mathbb{E}_{q(\neq z_i)} \left[z_i^T z_i + \frac{1}{\sigma^2} (W z_i + \mu - x_i)^T (W z_i + \mu - x_i) \right] =$$

$$= -\frac{1}{2} \left[z_i^T \left(I + \frac{1}{\sigma^2} \mathbb{E}_W W^T W\right) z_i + \frac{2}{\sigma^2} z_i^T \mathbb{E}_W W^T (\mu - x_i) \right]$$

$$\Rightarrow q(z_i) = \mathcal{N}(z_i | m_i, C), \quad C = \left(\frac{1}{\sigma^2} \mathbb{E} W^T W + I\right)^{-1} \quad m_i = \frac{1}{\sigma^2} C\, \mathbb{E} W^T (x_i - \mu)$$

$$\log q(W) = \mathbb{E}_{q(\neq W)} \log p(X, Z, W|\mu, \alpha, \sigma) \propto \mathbb{E}_{q(\neq W)} \left(-\frac{1}{2\sigma^2} \sum_{i=1}^{m} (W z_i + \mu - x_i)^T (W z_i + \mu - x_i) - \frac{1}{2} \sum_{i=1}^{n} \alpha_i w_i^T w_i \right) \propto$$

$$\propto -\frac{1}{2\sigma^2} \sum_{i=1}^{m} \left[\mathbb{E}_{q(\neq W)} z_i^T W^T W z_i + 2 m_i^T W^T (\mu - x_i) \right] - \frac{1}{2} \sum_{i=1}^{n} \alpha_i w_i^T w_i = -\frac{1}{2\sigma^2} \sum_{i=1}^{m} \operatorname{tr}\left(W \mathbb{E}_{q(z_i)} z_i z_i^T W^T\right)$$

$$-\frac{1}{2} \sum_{i=1}^{n} \alpha_i w_i^T w_i - \frac{1}{\sigma^2} \sum_{i=1}^{m} \operatorname{tr}\left((\mu - x_i) m_i^T W^T\right) = -\frac{1}{2} \operatorname{tr}\left(W^T \frac{1}{\sigma^2} \sum_{i=1}^{m} \mathbb{E} z_i z_i^T W + W^T \operatorname{diag}(\alpha) W\right)$$

$$-\frac{1}{\sigma^2} \sum_{i=1}^{m} \operatorname{tr}\left((\mu - x_i) m_i^T W^T\right) = -\frac{1}{2} \operatorname{tr}\left(W^T \left[\sum_{i=1}^{m} \mathbb{E} z_i z_i^T + \operatorname{diag}(\alpha)\right] W\right) - \frac{1}{\sigma^2} \sum_{i=1}^{m} \operatorname{tr}\left((\mu - x_i) m_i^T W^T\right) =$$

$$= -\frac{1}{2} \left(\sum_{i=1}^{n} w_i^T B_i^{-1} w_i - 2 \sum_{i=1}^{n} w_i^T B_i^{-1} u_i \right) \qquad B_i = \left(\frac{1}{\sigma^2} \sum \mathbb{E} z_i z_i^T + \operatorname{diag}(\alpha)\right)^{-1}$$

$$\mu_i = B_i \left(\frac{1}{\sigma^2} \sum_{i=1}^{m} (x_i - \mu) m_i^T\right)_i$$

$\uparrow$ i-столбец

т.е. $q(w_i) = \mathcal{N}(w_i | u_i, B_i)$

M-шаг: $\mathbb{E}_{q(Z,W)} \log p(X, Z, W | \mu, \alpha, \sigma) = \Phi(\mu, \alpha, \sigma) = \frac{1}{2\sigma^2}\sum_{i=1}^{m} \mathbb{E}[Xz_i+\mu-x_i]^T[Xz_i+\mu-x_i] -$

$-\frac{1}{2}\sum_{i=1}^{n}\alpha_i \|W_i\|_2^2 - \frac{n\cdot m}{2}\log\sigma^2 + \frac{d}{2}\sum_{i=1}^{n}\log\alpha_i$

$$\frac{\partial\Phi}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^m \mathbb{E}(Wz_i+\mu-x_i) = 0 \to \mu^* = \frac{1}{m}\sum_{i=1}^m x_i - \frac{1}{m}\mathbb{E}W\sum_{i=1}^m \mathbb{E}z_i$$

$$\frac{\partial\Phi}{\partial\alpha_i} = -\frac{1}{2}\mathbb{E}w_i^T w_i + \frac{d}{2\alpha_i} = 0 \to \alpha_i^* = \frac{d}{\mathbb{E}\|w_i\|_2^2}$$

$$\frac{\partial\Phi}{\partial\sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^m \mathbb{E}\|Wz_i+\mu-x_i\|_2^2 - \frac{m\cdot n}{2\sigma^2} = 0 \to \sigma^{2*} = \frac{1}{n\cdot m}\sum_{i=1}^m\left[\mathbb{E}\|Wz_i\|_2^2 + \|\mu^*-x_i\|_2^2 + 2\mathbb{E}z_i^T\mathbb{E}W^T\cdot(\mu^*-x_i)\right]$$

$\mathbb{E}W = \left(\sum_{i=1}^m (x_i-\mu)m_i^T\right)\left(\sum_{i=1}^m \|z_i\|_2^2 + \sigma^2\operatorname{diag}(\alpha)\right)^{-1}$ $\quad \mathbb{E}\|w_i\|_2^2 = tr(B_i + u_i u_i^T)$

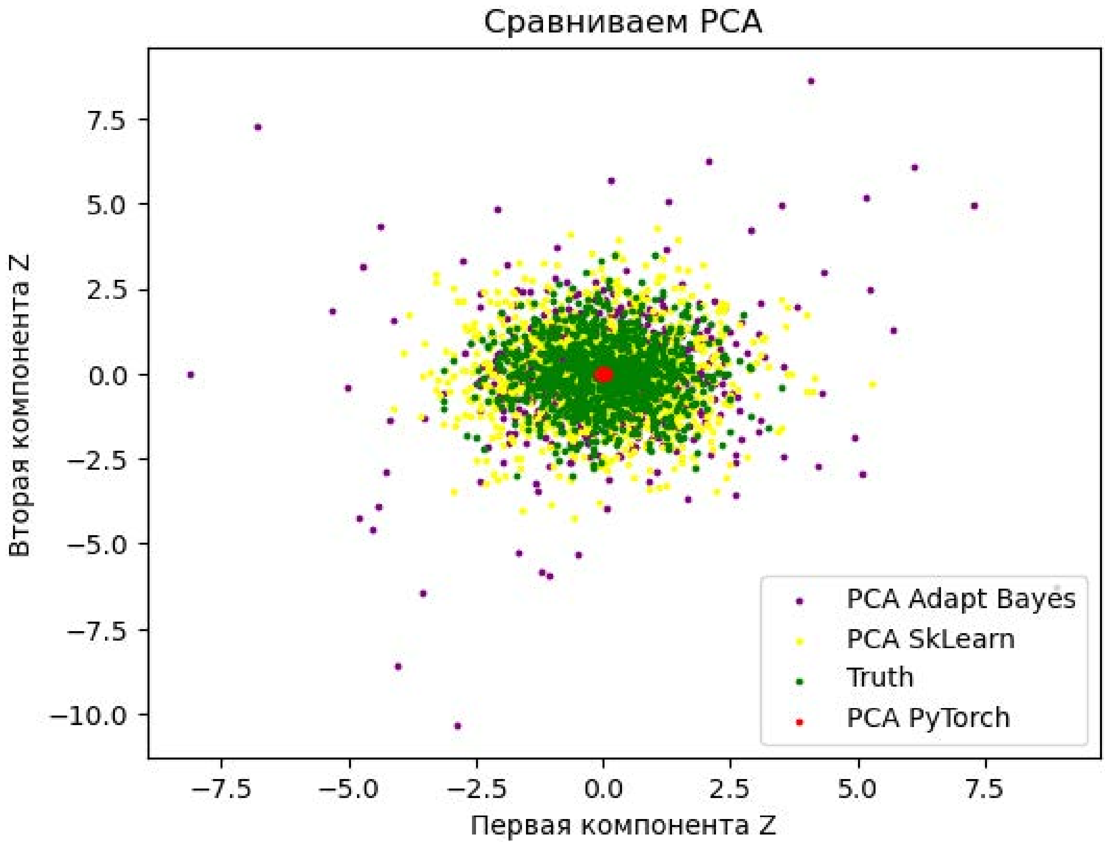
$\mathbb{E}z_i = m_i \quad \mathbb{E}\|z_i\|_2^2 = tr(A + m_i m_i^T)$

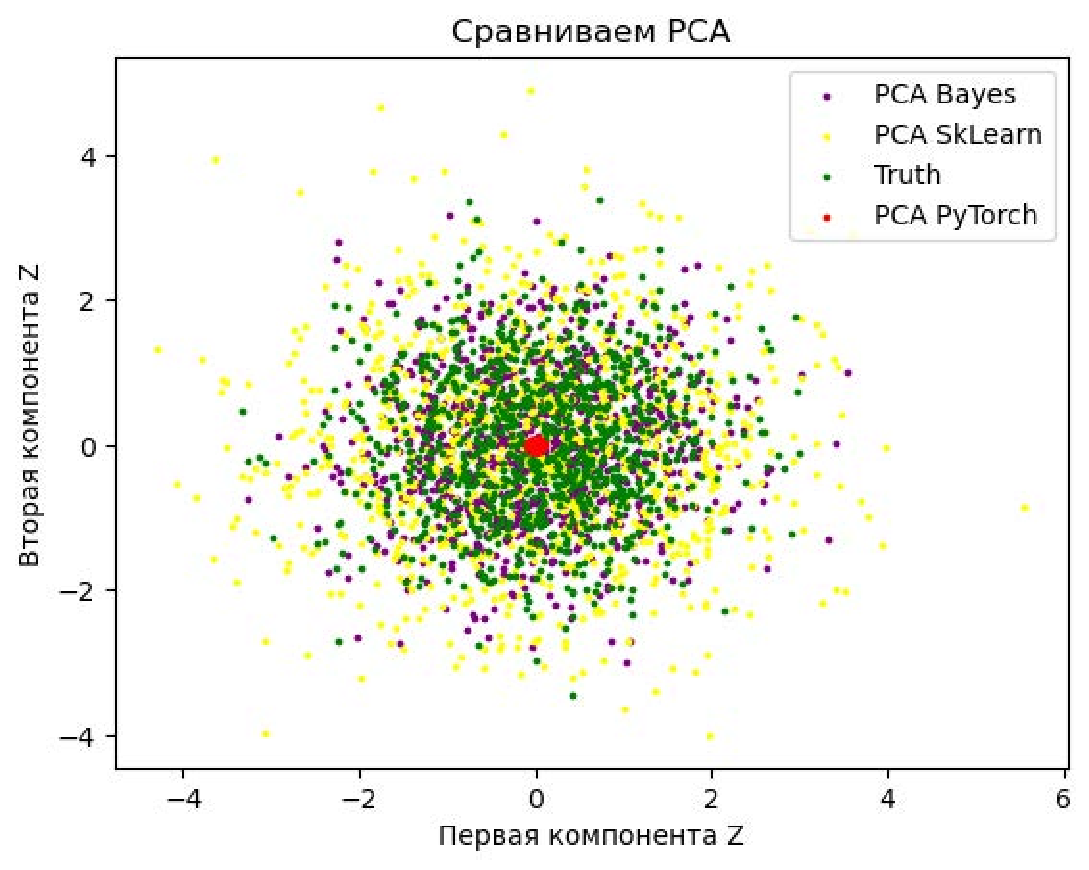
$\mathbb{E}W^TW = diag(\mathbb{E}\|w_i\|_2^2)$

```
Alpha =  tensor([12201.0068, 12201.0068, 12193.4219, 12201.0068,    33.8019, 12201.0068,
        12137.6543,  9274.9268, 12201.0068,  8050.2124])
```

+ Code + Markdown

Видно, что две альфы заметно меньше, чем остальные, поэтому можно сделать вывод, что автоподбор работает.
Получился автоподбор компонент работает похуже, чем Байесовский PCA, но в целом сопоставимо с sklearn версией и явно лучше, чем PyTorch версия, но его можно использовать, как подбор компонент, а потом уже Байесовский PCA с заданным количеством компонет.


Сравниваем PCA
Вторая компонента Z
Первая компонента Z
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
PCA Adapt Bayes
PCA SkLearn
Truth
PCA PyTorch


Сравниваем PCA
PCA Bayes
PCA SkLearn
Truth
PCA PyTorch
Вторая компонента Z
Первая компонента Z
4
2
0
−2
−4
−4
−2
0
2
4
6