1. (1) $\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)\right)=\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right)$

$=\sum_{s,a}p(s)\pi_0(s,a)\left(\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right)=\sum_{s,a}p(s)\pi_1(s,a)R(s,a)=\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}R(s,a)$.

(2) $\frac{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)\right)}{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}}=\frac{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right)}{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\frac{\pi_1(s,a)}{\pi_0(s,a)}}$

$=\frac{\sum_{s,a}p(s)\pi_0(s,a)\left(\frac{\pi_1(s,a)}{\pi_0(s,a)}R(s,a)\right)}{\sum_{s,a}p(s)\pi_0(s,a)\frac{\pi_1(s,a)}{\pi_0(s,a)}}=\frac{\sum_{s,a}p(s)\pi_1(s,a)R(s,a)}{\sum_{s,a}p(s)\pi_1(s,a)}$

$=\sum_{s,a}p(s)\pi_1(s,a)R(s,a)=\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}R(s,a)$

(3) 样本有限时 $\frac{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)\right)}{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}}=\frac{\sum_{i=1}^{n}\frac{\pi_1(s_i,a_i)}{\hat{\pi}_0(s_i,a_i)}p_iR(s_i,a_i)}{\sum_{i=1}^{n}\frac{\pi_1(s_i,a_i)}{\hat{\pi}_0(s_i,a_i)}p_i}$

对应 $n=1$ 时 $\frac{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)\right)}{\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}}=\frac{\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}R(s,a)}{\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}}=R(s,a)=\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}R(s,a)$

需估计量为 $\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}R(s,a)$，显然仅当 $\pi_0(s,a)\equiv\pi_1(s,a)$ 时二者相等，否则会产生偏差

(4) (i) $\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)+\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a)-\hat{R}(s,a))\right)$

$=\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)+\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\frac{\pi_1(s,a)}{\pi_0(s,a)}(R(s,a)-\hat{R}(s,a))$

$=\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}\hat{R}(s,a)-\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}(R(s,a)-\hat{R}(s,a))$

$=\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}R(s,a)$

(ii) $\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\left(\mathbb{E}_{a\sim\pi_1(s,a)}\hat{R}(s,a)+\frac{\pi_1(s,a)}{\hat{\pi}_0(s,a)}(R(s,a)-\hat{R}(s,a))\right)$

$=\mathbb{E}_{s\sim p(s),a\sim\pi_0(s,a)}\mathbb{E}_{a\sim\pi_1(s,a)}R(s,a)+0=\mathbb{E}_{s\sim p(s),a\sim\pi_1(s,a)}R(s,a)$

(5) (i) 应使用重要性采样估计量，由于IS不依赖对 $R(s,a)$ 建模，即使 $R(s,a)$ 复杂难以预测，IS仍可根据观测到数据估计.

(ii) 应使用回归估计量，此时避免了对 $\pi_0(s,a)$ 的依赖，减少行为策略建模影响。

2. $f_u(x)=\arg\min_{v\in V}\|x-v\|^2=u\arg\min_{\alpha\in\mathbb{R}}\|x-\alpha u\|^2=(x^Tu)u$，记 $U=\frac{1}{m}\sum_{i=1}^{m}x^{(i)}x^{(i)T}$ 为协方差矩阵，

则 $\arg\min_{\|u\|=1}\sum_{i=1}^{m}\|x^{(i)}-f_u(x^{(i)})\|_2^2=\arg\min_{\|u\|=1}\left(\sum_{i=1}^{m}\|x^{(i)}-(x^{(i)T}u)u\|_2^2=\sum_{i=1}^{m}\|x^{(i)}\|^2-2\sum_{i=1}^{m}\langle x^{(i)},(x^{(i)T}u)u\rangle+\sum_{i=1}^{m}\|x^{(i)T}u\|^2\right)$

$=\arg\min_{\|u\|=1}\left(\sum_{i=1}^{m}\|x^{(i)}\|^2-\sum_{i=1}^{m}(x^{(i)T}u)^2\right)=\arg\max_{\|u\|=1}\sum_{i=1}^{m}(x^{(i)T}u)^2=\arg\max_{\|u\|=1}\sum_{i=1}^{m}u^Tx^{(i)}x^{(i)T}u=\arg\max_{\|u\|=1}u^T\left(\sum_{i=1}^{m}x^{(i)}x^{(i)T}\right)u$

$=m\cdot\arg\max_{\|u\|=1}u^TSu=m\cdot\arg\max_{u\in\mathbb{S}^{m-1}}u^TSu$，$\mathbb{S}^{m-1}$ 为 $\mathbb{R}^m$ 中单位球面

由 $S=S^T$，作正交对角化 $P^TSP=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 且 $\lambda_1\geqslant\cdots\geqslant\lambda_n$，对 $\|u\|=1$，$\|Pu\|=1$ 且 P 是双射 $\Rightarrow P(\mathbb{S}^{m-1})=\mathbb{S}^{m-1}$

$\Rightarrow\arg\max_{u\in\mathbb{S}^{m-1}}u^TSu=\arg\max_{Pu\in\mathbb{S}^{m-1}}u^TP^T\mathrm{diag}(\lambda_1,\cdots,\lambda_n)Pu=\arg\max_{u=P^{-1}v\in\mathbb{S}^{m-1}}v^T\mathrm{diag}(\lambda_1,\cdots,\lambda_n)v=P^{-1}e_1$，对应特征值 λ_1 的特征方向

即为第一主成分

3. (1) $|B(V_1)(s)-B(V_2)(s)|=\left|R(s)+\gamma\max_{a\in A}\sum_{s'\in S}P_{sa}(s')V_1(s')-R(s)-\gamma\max_{a\in A}\sum_{s'\in S}P_{sa}(s')V_2(s')\right|$

$\leqslant\gamma\max_{a\in A}\left|\sum_{s'\in S}P_{sa}(s')(V_1(s')-V_2(s'))\right|\leqslant\gamma\max_{a\in A}\sum_{s'}P_{sa}(s')|V_1(s')-V_2(s')|$

$\leqslant\gamma\max_{a\in A}\sum_{s'}P_{sa}(s)\|V_1-V_2\|_\infty=\gamma\|V_1-V_2\|_\infty,\quad\forall s\in S$

$\Rightarrow\|B(V_1)-B(V_2)\|_\infty=\max_{s\in S}|B(V_1)(s)-B(V_2)(s)|\leqslant\gamma\|V_1-V_2\|_\infty$

(2) B 是 ∞ 范数诱导度量下压缩映射，若 $B(V)$ 有不唯一不动点 $x_1\neq x_2$，则

则 $\|x_1-x_2\|=\|B(x_1)-B(x_2)\|\leqslant\gamma\|x_1-x_2\|\Rightarrow\|x_1-x_2\|=0\Rightarrow x_1=x_2$ 矛盾，故不动点至多有一个.

4 (1) Lagrange 函数：$L(w,\alpha)=\frac{1}{2}w^Tw-\sum_{i=1}^{n}\alpha_i(w^Tx^{(i)}-1)$，$\nabla_wL=w-\sum_{i=1}^{n}\alpha_ix^{(i)}=0\Rightarrow w=\sum_{i=1}^{n}\alpha_ix^{(i)}$

此时 $\mathcal{L}(w,\alpha)=\frac{1}{2}\left(\sum_{i=1}^{n}\alpha_ix^{(i)}\right)^T\left(\sum_{i=1}^{n}\alpha_ix^{(i)}\right)-\sum_{i=1}^{n}\alpha_i\left(\left(\sum_{j=1}^{n}\alpha_jx^{(j)}\right)^T-1\right)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jx^{(i)T}x^{(j)}$

$\Rightarrow$ DP 为 $\max_\alpha\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jx^{(i)T}x^{(j)}$ s.t. $\alpha_i\geqslant0,\ \forall1\leqslant i\leqslant m$. or $\max_\alpha\mathbb{1}^T\alpha-\frac{1}{2}\alpha^TXX^T\alpha$ s.t. $\alpha\geqslant0$, $X=(x^{(1)},\cdots,x^{(m)})$

(2) 将内积 $x^{(i)T}x^{(j)}$ 替换为核函数 $k(x^{(i)},x^{(j)})$，由 Mercer / Hilbert-Schmidt，$\exists\,\phi(x)\in\mathcal{L}(V)$ s.t. $k(x^{(i)},x^{(j)})=\phi(x^{(i)})^T\phi(x^{(j)})$

此时问题转化为 $\max_\alpha\mathbb{1}^T\alpha-\frac{1}{2}\alpha^T\phi(X)\phi(X)^T\alpha$ s.t. $\alpha\geqslant0$，其中 $\phi(X)=(\phi(x^{(1)}),\cdots,\phi(x^{(m)}))$ 仅依赖 $k(x^{(i)},x^{(j)})$

而不依赖数据 $x^{(i)}$，故可以核化.

(3) 选择 α_i,α_j 进行优化，记 $k(x^{(i)},x^{(j)})=k_{ij}$，假定 $\alpha_i+\alpha_j=p$

$(\alpha_i,\alpha_j)=\arg\max_{\alpha_i,\alpha_j}\left(\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_jx^{(i)T}x^{(j)}\right)=\arg\max_{\alpha_i,\alpha_j}(\alpha_i+\alpha_j-\frac{1}{2}\alpha_i^2k_{ii}-\alpha_j^2k_{jj}-\alpha_i\alpha_jk_{ij})$

$\Rightarrow\alpha_i=\arg\max_{\alpha_i}\left(p-\frac{1}{2}k_{ii}\alpha_i^2-\frac{1}{2}k_{jj}(p-\alpha_i)^2-k_{ij}\alpha_i(p-\alpha_i)\right)=\arg\max_{\alpha_i}\left((-\frac{1}{2}k_{ii}+k_{ij}-\frac{1}{2}k_{jj})\alpha_i^2+p(k_{jj}-k_{ij})\alpha_i+p\right)$

$=-\frac{p(k_{jj}-k_{ij})}{2(-\frac{1}{2}k_{ii}+k_{ij}-\frac{1}{2}k_{jj})}=\frac{p(k_{jj}-k_{ij})}{k_{ii}-2k_{ij}+k_{jj}}$，$\alpha_j=p-\alpha_i=\frac{p(k_{ii}-k_{ij})}{k_{ii}-2k_{ij}+k_{jj}}$，只需更新这2个变量

可以通过选取不同变量对逐步逼近全局最优解.