PCA

PCA (empirical) (max using lagrange)

$$\sigma_v^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\pi(x_i-g))^T(\pi(x_i-g))$$

$$= \frac{1}{(n-1)}\sum_{i=1}^{n}(\mathbf{x}_i-\mathbf{g})^T\mathbf{v}\mathbf{v}^T(\mathbf{x}_i-\mathbf{g}) = \mathbf{v}^T\Sigma\mathbf{v}$$

$\max_{\mathbf{v}}(\mathbf{v}^T\Sigma\mathbf{v})$ subject to $(\mathbf{v}^T\mathbf{v}=1)$ and $(\mathbf{v}^T\mathbf{v}_1=0)$

$\pounds = \mathbf{v}^T\Sigma\mathbf{v} + \lambda(1-\mathbf{v}^T\mathbf{v})$ $\quad \partial_{\mathbf{v}}\pounds = 0$ $\quad \Sigma\mathbf{v} = \lambda\mathbf{v}$

$\pi = vv^T \quad \mathbf{v}^T\mathbf{v} = 1$ $\quad z_1 = \mathbf{v}^T(\mathbf{x}-\mathbf{g})$

eigenvalues = variance proj. $\quad \tilde{z}_{x1} = u_{x1}^T(x_k - g_x)$

PCA (mathematical) (max using lagrange)

$$var(y) = E[(y-E[y])^2]$$

$$E[y] = E[v^Tx] = v^TE[x] = v^Tg$$

$$var(y) = v^T E[(x-g)(x-g)^T]v$$

$$max_v(v^TSv) \ subject \ to \ (v^Tv=1)$$

$$\hat{S} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-g)(x_i-g)^T \approx \Sigma$$

$\mathrm{tr}(\Sigma) = \sum_{i=1}^{n}\lambda_i = \sigma^2,$

$$\text{Variance expliquée} = \frac{\sum_{j=1}^{k}\lambda_j}{\sum_{i=1}^{n}\lambda_i} = \frac{\sum_{j=1}^{k}\lambda_j}{\sigma_v^2}$$

la longueur de la projection de x sur v,

et la coordonnée de x dans la direction de cette composante principale.

$\sigma_y^2 = \mathbb{E}[(y-\mathbb{E}[y])^2] = \mathbb{E}[(v^Tx - v^Tg)^2]$

$cov(z_x, z_y) = \mathbb{E}[\tilde{z_x}\tilde{z_y}] = \mathbb{E}[u_x^T\tilde{x}\cdot u_y^T\tilde{y}] = u_x^T\,\mathbb{E}[\tilde{x}\tilde{y}^T]\,u_y = u_x^TS_{xy}u_y,$

FP

Équité au niveau des groupes:
* P(Yd = 1 |Z = 1) = P(Yd = 1 |Z = 0)
* E[Y |Z = 1] = E[Y |Z = 0]
* | P(Yd = 1 |Z = 1) – P(Yd = 1 |Z = 0) | ≤ ϵ
* | P(Yd = 1 |X, Z = 1) – P(Yd = 1 |X, Z = 0) | ≤ ϵ
* P(Yd = 1 |Y = 1,Z = 1) = P(Yd = 1 |Y = 1,Z = 0)
* P(Yd = 1 |Y = 1,Z = 1) = P(Yd = 1 |Y = 1,Z = 0) et P(Yd = 1 |Y = 0,Z = 1) = P(Yd = 1 |Y = 0,Z = 0)

Équité individuelle:
distY(hat(y_i), hat(y_j)) ≤ ϵ * distX(x_i, x_j)

Équité basée sur la causalité:
-> Déterminer si la variable protégée a un effet causal sur la décision et éliminer l'effet.

$$\frac{1}{n}|z^THXw| \le \epsilon$$

$$\min_w\left\{-\sum_{i=1}^{n}(y_i\log\hat{y}_i + (1-y_i)\log(1-\hat{y}_i))\right\}$$

cov(Z, hat(Y)) propr to E[Y |Z = 1] – E[Y |Z = 0]
1. Pré-traitement : See Other Face
2. In-training : Régularisation qui pénalise le lien entre Y_d et Z ou ajout de contrainte comme

FS

Max-relevance (highest corr. target)
Min-redundancy (least inter.corr. features)
MmRMR (max-rel. and min-red.)
Stepwise (add/rem. based on model perf.)
L1 reg. (make coeff. matrix sparse)

DA

Decompose variance

$$\sigma^2 = \frac{1}{n}\sum_{k=1}^{q=classes}SS(k) = \sigma_{(w)}^2 + \sigma_{(b)}^2$$

$Bv = (\frac{v^TBv}{v^T\Sigma v})\Sigma v$ $\quad \sigma_v^2 = v^T\Sigma v$

$SS_w(k) = \sum_{i\in C(k)}\|x_i - g(k)\|^2$

$2(v^T\Sigma v)Bv - 2(v^TBv)\Sigma v = 0$

$SS_b(k) = n(k)\|g(k)-g\|^2$ $\quad \Sigma^{-1}Bv = \lambda v$

$$\begin{cases}\sigma_{v(w)}^2 &= v^TWv\\ W_k &= \frac{1}{n(k)}\sum_{i\in C(k)}(x_i-g(k))(x_i-g(k))^T\\ W &= \frac{1}{n}\sum_{k=1}^{q}n(k)W_k\end{cases}$$

$x \mid y = k \sim \mathcal{N}(\mu_k, \Sigma)$

$P(y=k\mid x) \propto P(x\mid y=k)P(y=k)$

$\sigma_{v(b)}^2 = v^TBv$

$$max_v(\frac{\sigma_{v(b)}^2}{\sigma_v^2}) = max_v(\frac{v^TBv}{v^T\Sigma v}) \Rightarrow \partial v \quad B = [\sum^{q} n(k)\frac{(g(k)-g)(g(k)-g)^T}{n}$$

$$SS(k) = \sum_{i\in C(k)}(x_i-g)^T(x_i-g)$$
$$= \sum_{i\in C(k)}[(x_i-g(k)) + (g(k)-g)]^T[(x_i-g(k)) + (g(k)-g)]$$
$$= \sum_{i\in C(k)}(\|x_i-g(k)\|^2 + \|g(k)-g\|^2 + 2(x_i-g(k))^T(g(k)-g))$$

On assigne alors l'observation à la classe k pour laquelle $\delta_k(x)$ est maximale. Cette séparation est **linéaire** (hyperplan) car Σ est identique pour toutes les classes.

Intuition : même dans l'espace projeté, les observations de chaque classe forment des "nuages" gaussiens centrés autour de leur barycentre. Pour classifier une nouvelle observation, on ne regarde pas simplement sa distance euclidienne aux centres, mais sa **distance de Mahalanobis**, qui prend en compte la dispersion des données (via Σ).

MDS

$$d_{ij} = (s_{ii} - 2s_{ij} + s_{jj})^{1/2} \qquad K = XX^T$$

D>K $\begin{cases}K = -\frac{1}{2}HDH\\ \text{with K centered}\\ H = I - \frac{ee^T}{n}\end{cases}$ K>D $\begin{aligned}d_{ij}^2 &= \|x_i - x_j\|^2\\ &= \|x_i\|^2 + \|x_j\|^2 - 2x_i^Tx_j\\ &= k_{ii} + k_{jj} - 2k_{ij}\end{aligned}$

$$-\frac{1}{2}HDH = -\frac{1}{2}H(diag(K)\times e^T + e\times diag(K)^T - 2K)H$$
$$= -\frac{1}{2}(Hdiag(K)\times e^TH + Hediag(K)^TH - 2HKH)$$
$$= -\frac{1}{2}(0+0-2HKH) \qquad Ke = 0 \quad \text{et} \quad e^TK = 0$$
$$= K$$

$$\begin{cases}K = U\Lambda U^T = (U\Lambda^{1/2})(U\Lambda^{1/2})^T\\ U \text{ is the matrix of eigenvectors}\\ \Lambda \text{ is the diagonal matrix of eigenvalues}\end{cases}$$

$X = U\Lambda^{1/2}$ + eigenvalues (Λ>=0) ⇒ K positive semi-def ⇒ embedding in Euclidian space !

$X^TXu_1 = \lambda_1u_1$

i-th eigenvector of K = i-th coord. of individual on PCA axis $\quad XX^T(Xu_1) = \lambda_1(Xu_1)$

CCA

CCA (max using lagrange) $\quad z_x = u_x^Tx$ et $z_y = u_y^Ty$

$$\max_{u_x,u_y}\mathrm{cov}(z_x, z_y)\ \mathrm{var}(z_{x/y}) = 1$$

$S_{yx} = E[(y-g_y)(x-g_x)^T]$ COV

$S_{xx} = E[(x-g_x)(x-g_x)^T]$ VAR x

$S_{yy} = E[(y-g_y)(y-g_y)^T]$ " y

$$\begin{cases}\Sigma_{xy} = \Sigma_{yx} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k-g_x)(y_k-g_y)^T\\ \Sigma_{xx} = \frac{1}{n-1}\sum_{k=1}^{n}(x_k-g_x)(x_k-g_x)^T\\ \Sigma_{yy} = \frac{1}{n-1}\sum_{k=1}^{n}(y_k-g_y)(y_k-g_y)^T\end{cases}$$

$\partial_{u_x}\mathcal{L} = 0 \quad \Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}u_x = \lambda u_x$

$\partial_{u_y}\mathcal{L} = 0 \quad \Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}u_y = \lambda u_y$

$u_x = \frac{1}{\lambda_x}\Sigma_{xx}^{-1}\Sigma_{xy}u_y \quad \max_{u_x,u_y}\{u_x^T\Sigma_{xy}u_y\}$

$u_y = \frac{1}{\lambda_y}\Sigma_{yy}^{-1}\Sigma_{yx}u_x \quad u_x^T\Sigma_{xx}u_x = 1, \quad u_y^T\Sigma_{yy}u_y = 1.$

$\lambda_i = cov^2(z_{x,i}, z_{y,i}).$

MCA

$$\sum_{i=1}^{p}\sum_{j=1}^{p}cov(y_i,y_j) = E[(\sum_{i=1}^{p}y_i)^2] = u^TFu \qquad f_{kl} = \frac{n_{kl}}{n}$$

$$\sum_{i=1}^{p}cov(y_i,y_i) = \sum_{i=1}^{p}\sigma_{y_i}^2 = p. = \sum_{i=1}^{p}u_i^TD_iu_i = u^TDu, \qquad \lambda = \frac{u^TFu}{p} > 0$$

$\begin{cases}F = E[xx^T]\\ x = (x_1, x_2, ..., x_p)^T\\ u = (u_1, u_2, ..., u_p)^T\end{cases}$ $\quad \mathcal{L} = u^TFu + \lambda(p - u^TDu)$, $\quad Fu - \lambda Du = 0$, $\quad D^{-1}Fu = \lambda u$

F = Tableau de co-occurrence (ou contingence normalisée), f_kl est la fréquences de co-occurrence des modalité k et l (P(X_K=1, X_l=1)). D_i est diagonale car un attribut a qu'une modalité. D est bloc-diagonale -> contient les fréquences marginales des modalités. Lambda est la moyenne des covariances entre les variables. Vecteur u_i contient le score attribué à chaque modalité pour l'axe i.

DP 1

$$D*(s_k) = \min_{s_{k+1},\dots,s_N}[\sum_{i=k+1}^{N}d(s_i|s_{i-1})]$$
$$= \min_{s_{k+1},\dots,s_N}[d(s_{k+1}|s_k) + \sum_{i=k+2}^{N}d(s_i|s_{i-1})]$$
$$= \min_{s_{k+1}}[\min_{s_{k+2},\dots,s_N}[d(s_{k+1}|s_k) + \sum_{i=k+2}^{N}d(s_i|s_{i-1})]]$$
$$= \min_{s_{k+1}}[d(s_{k+1}|s_k) + \min_{s_{k+2},\dots,s_N}[\sum_{i=k+2}^{N}d(s_i|s_{i-1})]]$$
$$= \min_{s_{k+1}}[d(s_{k+1}|s_k) + D*(s_{k+1})]$$

N levels, s_k = state containing the state at level k

$$d(s_k = j|s_{k-1} = i)$$

MDP

MDP - Value Iteration

$$V^*(k_0) = \min_{(a_0,a_1,\dots)}\left[E_{s_1,s_2,\dots}\left[\sum_{t=0}^{\infty}c(u(s_t)|s_t)|s_0 = k_0, \pi\right]\right]$$

Separate first action from other (cost also)
Separate into two min (min a0) (min other a.) $\quad \mathbb{E}[z(t)] = \frac{1}{t}z(t) + \frac{1}{t}\sum_{\tau=1}^{t-1}z(\tau)$

$$V^*(k_0) = \min_{a_0\in U(k_0)}\left[\sum_{k_1}p(k_1|k_0,a_0)\times[c(a_0|k_0) + V^*(k_1)]\right]$$

$$\begin{cases}\hat{V}(k) \longleftarrow \min_{a\in U(k)}\left[c(a|k) + \sum_{k'=1}^{n}p(k'|k,a)\hat{V}(k')\right], k \ne d\\ \hat{V}(d) \longleftarrow 0 \text{ , where } d \text{ is the goal state}\end{cases}$$

$$argmin_{a\in U(k)}\left\{\sum_{k'}p(k'\mid k,a)\left[c(a\mid k) + \hat{V}(k')\right]\right\} \quad V^*(k) = \min_{a\in U(k)}\{Q^*(k,a)\}$$

$$Q_\pi(k,a) = c(a|k) + \sum_{k'=1}p(k'|k,a)V_\pi(k') \qquad = \mathbb{E}[z(t-1)] + \frac{1}{t}(z(t) - \mathbb{E}[z(t-1)])$$

$$\hat{Q}(k,a) \longleftarrow \hat{Q}(k,a) + \alpha(t)\left[\left[c(a|k) + \min_{a'\in U(k')}Q(k',a')\right] - Q(k,a)\right]$$

$$\begin{cases}\hat{Q}(k,a) \longleftarrow c(a|k) + \sum_{k'}p(k'|k,a)\min_{a'\in U(k')}\hat{Q}(k',a') \text{ with } k\ne d\\ \hat{Q}(d,a) = 0 \quad \hat{Q}(k,a) \approx f(x(k,a),w)\end{cases}$$

$$w_i \leftarrow w_i + \alpha(t)\left(c(a\mid k) + \min_{a'\in U(k')}\left\{\hat{Q}(k',a')\right\} - \hat{Q}(k,a)\right)\cdot\frac{\partial f(x(k,a),w)}{\partial w_i}$$

IR

$$Precision = \frac{TP}{TP+FP} \qquad Recall = \frac{TP}{TP+FN}$$

$$F-measure = \frac{2\times Precision\times Recall}{Precision + Recall}$$

$$P(R=1|d=x,u_k) = \frac{P(d=x|R=1,u_k)P(R=1|u_k)}{P(d=x|u_k)}$$

conditional independence between the words

$$\lambda = \frac{P(R=1|d=x,u_k)}{P(R=0|d=x,u_k)} = \frac{P(d=x|R=1,u_k)}{P(d=x|R=0,u_k)}\times\frac{P(R=1|u_k)}{P(R=0|u_k)}$$
$$= \frac{\prod_{n=i}^{n_w}P(d_n=x_n|R=1,u_k)}{\prod_{n=i}^{n_w}P(d_n=x_n|R=0,u_k)}\times\frac{P(R=1|u_k)}{P(R=0|u_k)}$$
$$\lambda \propto \frac{\prod_{n=i}^{n_w}P(d_n=x_n|R=1,u_k)}{\prod_{n=i}^{n_w}P(d_n=x_n|R=0,u_k)}$$

Tokenisation
Suppression des mots vides
Racinisation (stemming)
Lemmatisation
Reconnaissance des entités nomm

Modèle de recherche binaire d'indépendance, la fréquence des mots n'est pas non plus pris en compte. Ca revient à une sorte de classifieur naif bayésien

$$idf_i = -\log_2(P(w_i)) \approx -\log_2(\frac{df(w_i)}{n_d}) \quad \Big| \quad tf_{ij} = \frac{f_{ij}}{\sum_{k=1}^{n_w}f_{kj}}$$

$\begin{cases}D = U\Sigma V^T\\ UU^T = I\\ VV^T = I\end{cases}$ $\quad D_m = U\Sigma_mV^T$ Frobenius norm $\quad ||D_m - D||_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n_d}(d_{m\,ij} - d_{ij})^2 \quad \Sigma = \begin{bmatrix}\sigma_1 & 0\\ 0 & \sigma_2\end{bmatrix}$

D est la matrice de documents (ligne = mots, col = doc), contenant fréquence des mots dans les documents. Rang matrice = nombre d'info indépendance (donc ici nombre de thème sémantique). U et V basse dimension et vecteurs latents. Nombre de dimensions = ? (1. problème). SVD Computation si bcp de doc (2. Problème). Tronque jusque rang(D) les deriere val sing à 0 -> Meilleure approx de D

PG

Téléportation, Personnalisation, Matrice de Google, nœud pendant, G dense, P creuse, G régulière (val propre 1, vecteur propre gauche de P)

$$x_i = \sum_{j=1}^{n}\frac{w_{ij}x_j}{\sum_{i=1}^{n}w_{ij}}$$

$x_i(k+1) = \sum_{j=1}^{n}\frac{w_{ij}}{\sum_{i=1}^{n}w_{ij}}x_j(k)$ $\quad G = \alpha P + (1-\alpha)ev^T$

$$\begin{cases}x_i(k+1) = \sum_{j=1}^{n}p_{ij}x_j(k) = P^Tx(k)\\ x(k) = (P^T)^kx(0)\end{cases}$$

Converges when x(k+1)=x(k) ⇒ x = P^T x
Sol = left eigenvector of P (eigenvalue=1)

$\begin{cases}x \leftarrow Mx\\ x \leftarrow \frac{x}{||x||}\end{cases}$ M = P^T $\quad$ TP $\quad G = \alpha P + (1-\alpha)\frac{ee^T}{n}$

$= \alpha x^T(k)P + \frac{1-\alpha}{n}e^T \qquad (I - \alpha P)^Tx = (1-\alpha)v$

HITS $\begin{cases}x^a = \eta W^Tx^h\\ x^h = \mu Wx^a\end{cases}$ $\begin{cases}x^a = \eta\mu W^TWx^a\\ x^h = \eta\mu WW^Tx^h\end{cases}$

D_{in} = diagonal matrix containing the indegree of each node
D_{out} = diagonal matrix containing the outdegree of each noc
C_{cit} = matrix containing the cocitation
C_{ref} = matrix containing the coreference
$W^TW = D_{in} + C_{cit}$
$WW^T = D_{out} + C_{ref}$

PCA non centré.

Méthode de la puissance pour PageRank

Les scores de PageRank correspondent à la **distribution stationnaire** de la chaîne de Markov, soit le **vecteur propre gauche** de P associé à la valeur propre 1 — ce qui revient à trouver le **vecteur propre droit** de P^T associé à cette même valeur propre.

Col Rec

K-NN

		Individual j 1	0	Totals
Individual i	1	a	b	$a+b$
	0	c	d	$c+d$
Totals		$a+c$	$b+d$	$p=a+b+c+d$

$KNN(p) = \{i_1, i_2, ..., i_k\}$ the cl

$$pred(p,i) = \frac{\sum_{q\in KNN(p)}sim(p,q)\cdot w_{qi}}{\sum_{q\in KNN(p)}sim(p,q)}$$

Sim cosinus ou celle ci: $\quad \frac{a}{a+b+c}$

ItemRank

$x_i(k+1) = \alpha P^Tx_i(k) + (1-\alpha)v_i$
v_i contains $1/n_i$ for each item user i bought, 0 otherwise
n_i is the number of bought items
P is the transition matrix derived from the graph
$(1-\alpha)v_i$ is the probability of being teleported on a bought item noc

$$x_i = (1-\alpha)(I - \alpha P^T)^{-1}v_i$$

u_i = tastes | v_j – attributes $\quad$ Nn Matrix Fact

$W \approx UV^T$ $\quad \begin{cases}\min_{U,V}||W - UV^T||_F^2\\ \text{subject to } U\ge 0, V\ge 0\end{cases}$

$\sum_{(i,j)\in\epsilon}(w_{ij} - u_i^Tv_k)^2$

Delete some value from W and use them as a test set, cross-validation, precision & recall

EM algorithm for items collaboration
x=individuals (m) | y=items (n) | z=class (l)

$$P(x,y) = \sum_{k=1}^{l}\overbrace{P(x|z=k)P(y|z=k)}^{x,y \text{ INDEPEDENT}}P(z=k)$$

$P(k) = P(z=k) \approx \hat{P}(k)$ class prior probabilities
$P(x|k) = P(x=x|z=k) \approx \hat{P}(x|k)$ within-class observation of users
$P(y|k) = P(y=y|z=k) \approx \hat{P}(y|k)$ within-class observation of items

$$P(y|x) = \frac{P(x,y)}{P(x)} = \frac{\sum_{k=1}^{l}P(x|k)P(y|k)P(k)}{\sum_{k'=1}^{l}P(x|k')P(k')}$$

1. estimate value of z for each user with parameters via log-likelihood
2. maximization step = re-estimate parameters

$$\widehat{P}(z) \leftarrow \frac{\sum_{x=1}^{m}\sum_{y=1}^{n}\widehat{P}(z|x,y)\,n(x,y)}{\sum_{x'=1}^{m}\sum_{y'=1}^{n}n(x',y')}$$

$$\widehat{P}(x|z) \leftarrow \frac{\sum_{y=1}^{n}\widehat{P}(z|x,y)\,n(x,y)}{\sum_{x'=1}^{m}\sum_{y'=1}^{n}\widehat{P}(z|x',y')\,n(x',y')}$$

$$\widehat{P}(y|z) \leftarrow \frac{\sum_{x=1}^{m}\widehat{P}(z|x,y)\,n(x,y)}{\sum_{x'=1}^{m}\sum_{y'=1}^{n}\widehat{P}(z|x',y')\,n(x',y')}$$

$$Q(\theta\mid\hat{\theta}^{(t)}) = \mathbb{E}_{z\sim P(z|x,y;\hat{\theta}^{(t)})}[\log P(x,y,z\mid\theta)]$$
$$\hat{\theta}^{(t+1)} = \arg\max_\theta Q(\theta\mid\hat{\theta}^{(t)})$$

$$\widehat{P}(z|x,y) \leftarrow \frac{\widehat{P}(x|z)\,\widehat{P}(y|z)\,\widehat{P}(z)}{\sum_{k=1}^{l}\widehat{P}(x|k)\,\widehat{P}(y|k)\,\widehat{P}(k)}$$

DP 2

$$edit\_dist(x,y) = D^*(x_{|x|}^{|x|}, y_0^{|y|}) \begin{cases}D^*(0,n-1) = 0\\ D^*(i,n-1) = \infty & \forall i\ne 0\end{cases}$$

$$D^*(i,k) = \min_{j\in Succ(i)}\{c_{ij} + D^*(j,k+1)\}$$

$$lcs(x,y) = \frac{1}{2}(|x| + |y| - edit\_dist(x,y))$$

$$\begin{cases}D^*(s_N) = 0\\ D^*(s_k = i) = \min_{s_{k+1}}\{d(s_{k+1}\mid s_k = i) + D^*(s_{k+1})\}\\ D^* = \min_{s_0}\{D^*(s_0)\}\end{cases}$$

$$\begin{cases}D^*(s_0) = 0\\ D^*(s_k = i) = \min_{s_{k-1}}\{d(s_k = i\mid s_{k-1}) + D^*(s_{k-1})\}\\ D^* = \min_{s_N}\{D^*(s_N)\}\end{cases}$$

$$\begin{cases}g(1,1) = d(r_1^n, o_1)\\ g(i,j) = \min\begin{cases}g(i-1,j) + d(r_i^n, o_j) & \text{(insertion)}\\ g(i-1,j-1) + 2d(r_i^n, o_j) & \text{(diagonale)}\\ g(i,j-1) + d(r_i^n, o_j) & \text{(suppression)}\end{cases}\end{cases}$$

$$D(R_n, O) = \frac{1}{I+J}\cdot g(I,J)$$

$$D^*(x_i^{|x|}, y_0^j) = \min\begin{cases}D^*(x_i^{|x|}, y_0^{j-1}) + 1 & \text{(insertion)}\\ D^*(x_{i-1}^{|x|}, y_0^j) + 1 & \text{(suppression)}\\ D^*(x_{i-1}^{|x|}, y_0^{j-1}) + \delta_{ij} & \text{(substitution)}\end{cases}$$

Monotonie : $i_k \ge i_{k-1}, \quad j_k \ge j_{k-1}$

Conditions aux bords : $i_1 = j_1 = 1$ et $i_K = I, \quad j_K = J$

Continuité : $i_k - i_{k-1} \le 1, \quad j_k - j_{k-1} \le 1$

$c_{ij} = \infty$ s'il n'existe pas de lien entre i et j

$c_{ij} > 0$ si un lien existe entre i et j, avec $i \ne 0$

$c_{0i} = \infty \quad \forall j \ne 0$ et $c_{00} = 0$: le noeud destination est absorbant

$c_{ii} = \infty \quad \forall i \ne 0$: les boucles (self-loop) ne sont pas autorisées

MC

$$P(s_{t+2} = j|s_t = i) = \sum_{k=1}^{n}P(s_{t+2} = j, s_{t+1} = k|s_t = i)$$

SUM BEFORE P!

$$P(s_{t+2} = j|s_t = i, s_{t+1} = k)P(s_{t+1} = k|s_t = i)$$
$$P(s_{t+2} = j|s_{t+1} = k)P(s_{t+1} = k|s_t = i)$$

$p_{kj}p_{ik} = [P^2]_{ij} \rightarrow P(s_{t+\tau} = j|s_t = i) = [P^\tau]_{ij}$

$$x(t) = P^Tx(t-1) = (P^T)^tx(0)$$

in state j at time t $x_j(t) = x(t)^Te_j = x(0)^TP^te_j$
in state=j at time=t given start state=i time=0

$x_j(t|s_0 = i) = x_{j|i}(t) = (e_i)^TP^te_j = (e_j)^T(P^T)^te_i$

$$x_i(t) = \mathbb{P}(s_t = i) = \sum_{i=1}^{n}\mathbb{P}(s_t = i, s_{t-1} = k) = \sum_{k=1}^{n}\mathbb{P}(s_t = i\mid s_{t-1} = k)\cdot\mathbb{P}(s_{t-1} = k) = \sum_{k=1}^{n}p_{ki}\cdot x_k(t-1)$$

Markov Chain (absorbing)

$$\mathbf{P} = \begin{array}{c|cc} & \text{TR.} & \text{ABS.}\\ \hline \text{TR.} & \mathbf{Q} & \mathbf{R}\\ \text{ABS.} & \mathbf{0} & \mathbf{I}\end{array}$$

P^t : Q → 0 if t⇒inf
N = I + Q + Q^2 + ...
N = (I - Q)^{-1}

$n_{ij} = e_i^TNe_j = e_i^T\left(\sum_{t=0}^{\infty}Q^t\right)e_j = \sum_{t=0}^{\infty}e_i^TQ^te_j = \sum_{t=0}^{\infty}x_{j|i}(t)$ car,

n_{ij} = expected number visit through j sarting i
n = Ne ⇒ nb step until absorbed foreach start state

= proba absorbed by j starting from i

$$b_{ij} = \sum_{t=0}^{+\infty}\sum_{k=1,k\in TR}^{n_{tr}}x_{k|i}(t)r_{kj} = \sum_{k=1,k\in TR}^{n_{tr}}[\sum_{t=0}^{+\infty}e_i^TQ^te_k]r_{kj} = \sum_{k=1,k\in TR}^{n_{tr}}n_{ik}r_{kj} = [NR]_{ij}$$

$$\bar{m} = \sum_{t=0}^{\infty}\gamma^t\sum_{i=1}^{n}x_i(t)m_i = \sum_{t=0}^{\infty}\gamma^tm^T(P^T)^tx(0) = \sum_{t=0}^{\infty}\gamma^tx^T(0)P^tm = x^T(0)\left(\sum_{t=0}^{\infty}\gamma^tP^t\right)m = x^T(0)(I-\gamma P)^{-1}m$$

Markov Chain (appli)
n clusters = states
n-th cluster absorbing (losing customer)
P estimated from monthly moves
m_i = average profit per segment of customer
RFM = recency/frequency/monetary value

HMM

n hidden states = s(t) = {1, ..., n}
p observations = x(t)={o_1, ..., o_p} ⇒ [x_1, ..., x_T]
$\Pi = \{\pi_i\}$ initial proba; P(s(1)=i) P=st. trans. proba
$B = \{b_i(o_k)\}$ emission proba $P(x(t)=o_k \mid s(t)=i)$

$\mathbf{x} = [x_1,\dots,x_T]^T$, $\theta = \{\Pi, P, B\}$
Compute $P(\mathbf{x}\mid\theta)$

$$P(\mathbf{x}\mid\theta) = \sum_{(s_1\dots s_T)}\pi_{s_1}b_{s_1}(x_1)\prod_{t=1}^{T-1}p_{s_ts_{t+1}}b_{s_{t+1}}(x_{t+1})$$

Forward procedure:

$$\alpha_i(t) = P(x_1,\dots,x_t, s_t = i|\theta)$$
$$\begin{cases}\alpha_j(1) = b_j(x_1)\pi_j\\ \alpha_j(t+1) = \left(\sum_{i=1}^{n}p_{ij}\alpha_i(t)\right)b_j(x_{t+1})\end{cases}$$
$$P(\mathbf{x}|\theta) = \sum_{j=1}^{n}\alpha_j(T)$$

Backward procedure:

$$\beta_i(t) = P(x_{t+1},\dots,x_T|s_t = i,\theta)$$
$$\begin{cases}\beta_i(T) = 1\\ \beta_i(t) = \sum_{j=1}^{n}p_{ij}b_j(x_{t+1})\beta_j(t+1)\end{cases}$$
$$P(\mathbf{x}|\theta) = \sum_{i=1}^{n}\pi_i\beta_i(1) \qquad P(x|\theta) = \sum_{i=1}^{n}\alpha_i(t)\beta_i(t)$$

Post-Processing:
* échanger les individus protégés rejetés les plus proches de la frontière de décision vers acceptés
* échanger les individus non protégés acceptés les plus proches de la frontière vers rejetés
* continuer jusqu'à ce que la propriété 3 ci-dessus soit réspectés
* -> Méthode non optimale

$$= \min_{(a_0,a_1,\dots)}\left\{\sum_{k_1,k_2,\dots}P(s_1 = k_1, s_2 = k_2,\dots\mid s_0 = k_0,\pi)\cdot\sum_{t=0}^{\infty}c(u(k_t)\mid k_t)\right\}$$
$$= \min_{(a_0,a_1,\dots)}\left\{\sum_{k_1,k_2,\dots}P(s_2 = k_2, s_3 = k_3,\dots\mid s_1 = k_1,\pi)\cdot P(s_1 = k_1\mid s_0 = k_0, u(k_0) = a_0)\cdot\left[c(a_0\mid k_0) + \sum_{t=1}^{\infty}c(u(k_t)\mid k_t)\right]\right\}$$
$$= \min_{a_0}\left\{\sum_{k_1}P(s_1 = k_1\mid s_0 = k_0, u(k_0) = a_0)\cdot\left[c(a_0\mid k_0) + \min_{(a_1,a_2,\dots)}\left\{\sum_{k_2,k_3,\dots}P(s_2 = k_2,\dots\mid s_1 = k_1,\pi)\cdot\sum_{t=1}^{\infty}c(u(k_t)\mid k_t)\right\}\right]\right\}$$

$$b_{ij}(t) = \sum_{t=0}^{\infty}\mathrm{P}(s_t\in\mathrm{TR}, s_{t+1} = j|s_0 = i) = \sum_{t=0}^{\infty}\sum_{k\in\mathrm{TR}}\mathrm{P}(s_t = k, s_{t+1} = j|s_0 = i)$$

$P = (p_{ij})$ avec $p_{ij} = \mathbb{P}(s(t+1) = j\mid s(t) = i)$

Viterbi Algorithm (DP Algo)

On modélise chaque mot $w^{(k)}$ avec un HMM $\theta^{(k)}$.

$$\delta_j(1) = \pi_jb_j(x_1)$$
$$\delta_j(t+1) = \max_i\{\delta_i(t)\cdot p_{ij}\cdot b_j(x_{t+1})\}$$
$$\psi_j(t+1) = \arg\max_i\{\delta_i(t)\cdot p_{ij}\cdot b_j(x_{t+1})\} \qquad \hat{k} = \arg\max_k\left(\max_j\delta_j^{(k)}(T)\right)$$

$cov(Z,\hat{Y}) \propto \mathbb{E}[\hat{Y}\mid Z=1] - \mathbb{E}[\hat{Y}\mid Z=0]$
$x_{ik} = w_1z_{i1} + w_2z_{i2} + \dots$
$e_{ik} = x_{ik} - \tilde{x}_{ik}$ is the residual
$\Rightarrow x_{ik} = w_1z_{i1} + w_2z_{i2} + \dots + e_{ik}$

$$\left|\frac{1}{n-1}\sum_i\hat{y}_{ik}\tilde{z}_i\right| \le \epsilon \qquad \max_{\hat{y}_{ik}}\left\{-\sum_{i,k}\hat{y}_{ik}\log\hat{y}_{ik}\right\}$$