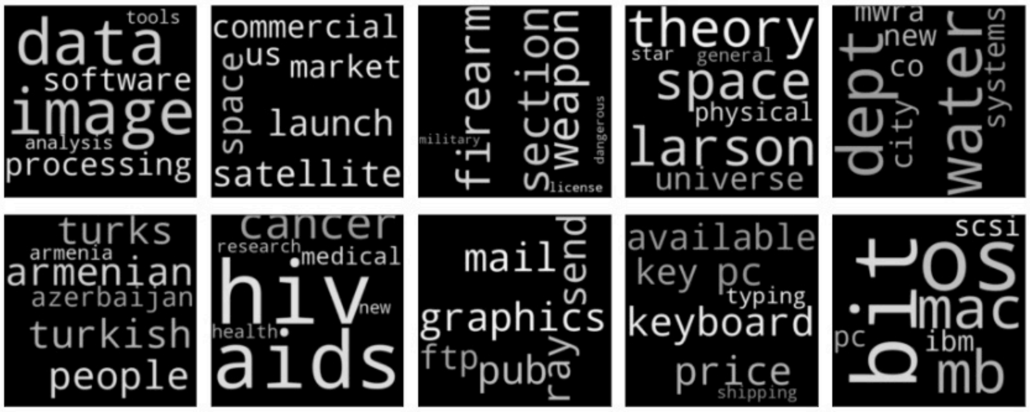
Excercise 1

w/ non-negative code AND dictionary, bag of words vectorizer.



w/ non-negative dictionary, unconstrained code, bag of words vectorizer

Between the first image and figure S2, we can see that the images are VERY similar and for the ones of the same catergory they are identical. The only difference between pictures w/ the same catergory is that some words are diffrent shades or size. This is because the frequency is scaled differently. Figure S2 uses a vectorizer that normalizes our word count frequency, while bag of words instead simply uses the word count in a catergory, thus resulting in the diffrent shades and sizes of words between both word clouds.

The second image is also similar too figure S3, it is not able to learn the catergories well but the word size and shading is different.

Exercise 2 | Let $x \in \mathbb{R}^d$ and $C \subseteq \mathbb{R}^d$. Now:

$$\text{proj}_C(x) := \underset{\mu \in C \subseteq \mathbb{R}^d}{\arg\min} \|x-\mu\|_F^2$$

i) $\|x-\mu\|_F^2 = tr((x-\mu)^T(x-\mu)) = tr(\mu^T\mu) - 2tr(x^T\mu) + tr(x^Tx)$

$$\frac{\partial}{\partial \mu_a} \|x-\mu\|_F^2 = \frac{\partial}{\partial \mu_a}(tr((x-\mu)^T(x-\mu)) = \frac{\partial}{\partial \mu_a}\left(\sum_i (x_i-\mu_i)^2\right)$$

$$= 2(x_a - \mu_a)(-1)$$

$$= 2(\mu - x)_a$$

$\blacksquare$

ii) Assume C is span$\langle v \rangle : \{av \mid a \in \mathbb{R}\}$, then:

$$\text{Proj}_v(x) := \text{Proj}_{\langle v \rangle}(x) = \underset{a \in \mathbb{R}}{\arg\min} \|x - av\|_F^2$$

Solving above directly gives us:

$$\frac{\partial}{\partial a_n}(\|x-av\|_F^2) = \frac{\partial}{\partial a_n}\left(\sum_i (x_i - av_i)^2\right)$$

$$= \frac{\partial}{\partial a_u} \sum_i \cancel{x_i^2} - 2ax_iv_i + a^2v_i^2$$

$$= \left(\sum_i -2x_iv_i + a^2v_i^2\right)_n$$

Now setting equal to zero and isolating a:

$$\sum_i -2x_iv_i + 2av_i^2 = 0$$

$$\sum_i \cancel{2}av_i^2 = \sum_i \cancel{2}x_iv_i$$

$$a = \frac{\sum_i x_iv_i}{\sum_i v_i^2} = \frac{x^Tv}{v^Tv} = \frac{x^Tv}{\|v\|^2}$$

$\|v\|^2 = 1$
v is normal

Now we have that $a = x^T v$ and plugging back into $u = av$ gives us:

$$\text{Proj}_{\langle u \rangle}(x) = av = (x^T v)v$$

iii) let $a = \begin{bmatrix} a_1 \\ \vdots \\ a_r \end{bmatrix}$ s.t. $C = \{ W a \in \mathbb{R}^d \mid a \in \mathbb{R}^r \}$ and $W = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}$

so now let $u = Wa$ s.t.:

$$\|x - u\|^2 = \sum_i \Big(x_i - \sum_j w_{ij} a_j\Big)^2$$

Now solving our optimization problem gives us:

$$\frac{\partial}{\partial a_j} = \sum_i 2\Big(x_i - \sum_j w_{ij} a_j\Big)(-w_{ij})$$

which we can then set to 0 and solve for j

$$\sum_i 2\Big(x_i - \sum_j w_{ij} a_j\Big)(-w_{ij}) = 0$$

$$\sum_i (x_i - (Wa)_i) w_{ij} = 0$$

$$\sum_i x_i w_{ij} - \sum_i (Wa)_i w_{ij} = 0$$

$$(W^T x)_j - (W^T W a)_j = 0$$

since W is orthonormal we know $W^T W = I$

$$W^T x = W^T W a$$

$$\frac{W^T x}{I} = a$$

$$a = W^T x$$

and finally plugging back into our initial equation of $u = Wa$ gives us:

$$\text{Proj}_{\langle W \rangle}(x) = Wa = W(W^T x) \quad \blacksquare$$

Excercise 3

Show that $tr(ABC) = tr(BCA)$

$$tr(ABC) = \sum_i (ABC)_{ii} = \sum_i \sum_j A_{ij} (BC)_{ji} = \sum_i \sum_j \sum_k A_{ij} B_{jk} C_{ki}$$

Scalers order does not matter

$$= \sum_i \sum_j \sum_k B_{jk} C_{ki} A_{ij}$$

$$= \sum_i \sum_j (BC)_{ji} A_{ij}$$

$$= \sum_j (BCA)_{jj}$$

$$= tr(BCA) \quad \blacksquare$$

Exercise 4

i) Show that all eigenvalues of A are real:

Let $Av = \lambda v$ s.t. $v \neq 0$ where $v \in \mathbb{C}^{n\times n}$, $A \in \mathbb{R}^{n\times n}$ and $\mathbb{R}^{n\times n} \subseteq \mathbb{C}^{n\times n}$,
now suppose we have:

$$\langle v, Av\rangle = v^*Av \rightarrow \text{s.t. } v^* = \overline{v^T}$$
$$= v^*\lambda v$$
$$= \lambda v^* v \qquad \text{(I)}$$
$$= \lambda \|v\|^2 \rightarrow v^*v = \left.\begin{matrix} v_1\bar{v}_1 \\ + \\ \vdots \\ v_1\bar{v}_n \end{matrix}\right\} \text{real}$$

Now again we have:

$$\langle v, Av\rangle = v^*Av$$
$$= (A^*v)^*v$$
$$= (\lambda v)^* v \rightarrow A^* = \overline{A^T} = \overline{A} = A$$

since A is symmetric A^TA

we know that A is real so $\overline{A} = A$

$$= \overline{\lambda v^T} v = \bar{\lambda}\|v\|^2$$

Now combining this result w/ what we got in (I), we have that: $\lambda\|v\|^2 = \bar{\lambda}\|v\|^2$

$$\lambda = \bar{\lambda}$$

and we can see that λ is $\bar{\lambda}$, so we conclude all λ's are real as our proof considered $v \in \mathbb{C}^{n\times n}$ arbitary.

ii)

Let $Av_1 = \lambda_1 v_1$, where $\lambda_1 \in \mathbb{R}$ by **i)** and $\|v_1\| = 1$ as v is orthonormal where Q is an orthonormal basis of $\mathbb{R}^n$

$Q^TQ = I$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = [\ \ ]$$

Now we can see:

$$Q^TAQ = Q^T\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix}\begin{bmatrix} | & | & & | \\ v_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

$$= Q^T \left[Av_1 \,\middle|\, Au_2 \,\middle|\, \cdots \,\middle|\, Au_n \right] = Q^T \left[\lambda v_1 \,\middle|\, Au_2 \,\middle|\, \cdots \,\middle|\, Au_n \right]$$

$$= \begin{bmatrix} - & v_1 & - \\ - & u_2 & - \\ & \vdots & \\ - & u_n & - \end{bmatrix} \left[\lambda v_1 \,\middle|\, Au_2 \,\middle|\, \cdots \,\middle|\, Au_n \right]$$

$$= \left[\begin{array}{c|ccc} \lambda v_1^T v_1 & v_1^T A u_2 & \cdots & v_1^T A u_n \\ \hline \lambda u_2 v_1 & & & \\ \vdots & & A' & \\ \lambda u_n v_n & & & \end{array}\right]$$

orthogonal

$v_1^T A u_j =$

$(A^T v_1)^T u_j =$

$(A v_1)^T u_j =$

$(\lambda_1 v_1)^T u_j = 0$

$u_j A u_j = A u_j^T u_j$

$= A$ since $u_j^T u_j = 1$

but we know top row and left rows are zeroed out

we are left w/ $A' \in \mathbb{R}^{(n-1) \times (n-1)}$

$v_1^T v_1 = 1$

$$= \left[\begin{array}{c|ccc} \lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right]$$