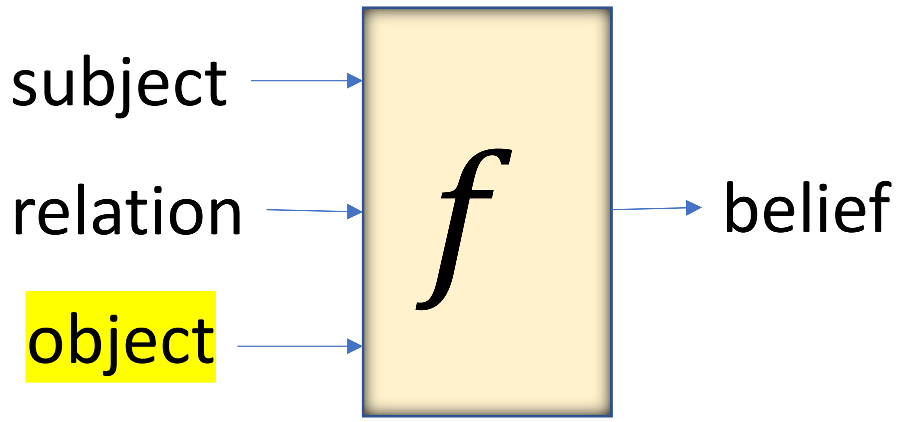
Knowledge and Retrieval

Translation and Rotation Models

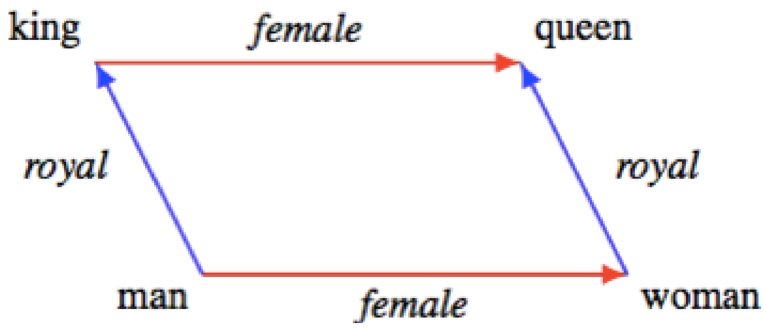
Design of scoring function f

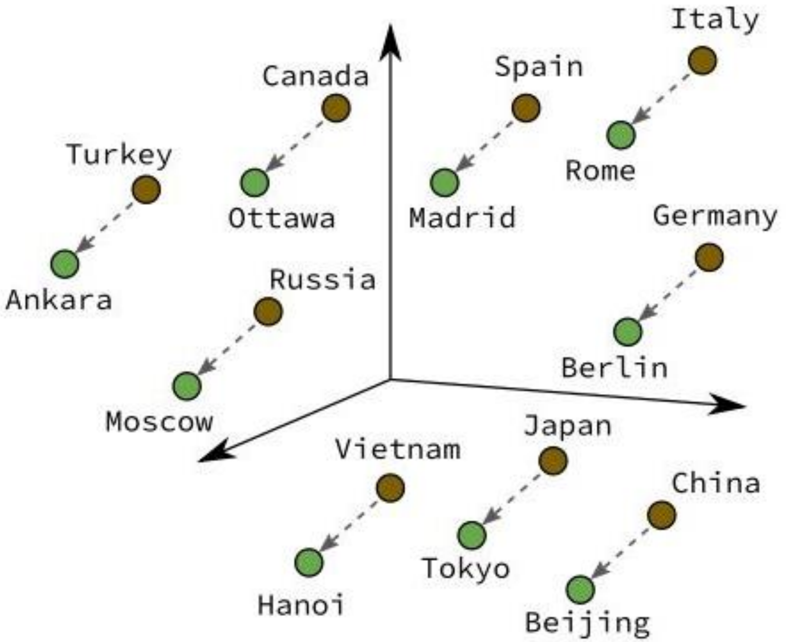
- Design a scoring function $f(s, r, o)$ for the belief in fact $(s, r, o) \in \mathrm{KG}$
- "In principle," deep networks are universal function approximators
- In practice, several years and hundreds of papers proposing f and associated loss function and sampling strategy



Quick review of word2vec and GloVE

- Factor document-word matrix to obtain dense vector embeddings of words
- Related to singular value decomposition
- Each (1-1) relation can be modeled as a distinct displacement or *translation*
- ⇒ "TransE"





Country-Capital

TransE

- $f(s, r, o) = \|\boldsymbol{s} + \boldsymbol{r} - \boldsymbol{o}\|$ where $\boldsymbol{s}, \boldsymbol{r}, \boldsymbol{o} \in \mathbb{R}^D$
- Subject, translated by relation, should be near object, otherwise low confidence that fact $(s, r, o) \in \mathrm{KG}$
- Low score $\Leftrightarrow$ high confidence and vice versa
- Loss adjusted to the hinge form

$$\frac{1}{K} \sum_k \max\{0, \mathrm{margin} \underbrace{- f(s'_k, r, o'_k)}_{\text{Want large}} + \underbrace{f(s, r, o)}_{\text{Want small}}\}$$

TransE benefits and limitations

- 👍Simple and fast
- 👍Very few hyperparameters (margin and negative samples per positive sample)
- 👎Cannot model 1-to-many, many-to-1, many-to-many relations
  - Obama + attended ≈ Occidental College
  - Obama + attended ≈ Harvard Law School
  - ⇒ Occidental College ≈ Harvard Law School
- 👎Cannot model symmetric relations
  - $(\boldsymbol{s}+\boldsymbol{r}=\boldsymbol{o})$ and $(\boldsymbol{o}+\boldsymbol{r}=\boldsymbol{s}) \Rightarrow \boldsymbol{r}=\mathbf{0}$
- Many patches “XtransY”

TransH: Early fix to TransE

- Represent r by *two* artifacts
  - Hyperplane with unit normal $\boldsymbol{p}_r$
  - Displacement $\boldsymbol{d}_r$ as before (was called $\boldsymbol{r}$)
  - Expect $(\boldsymbol{s} \downarrow \boldsymbol{p}_r) + \boldsymbol{d}_r \approx (\boldsymbol{o} \downarrow \boldsymbol{p}_r)$

Subject projected to hyperplane

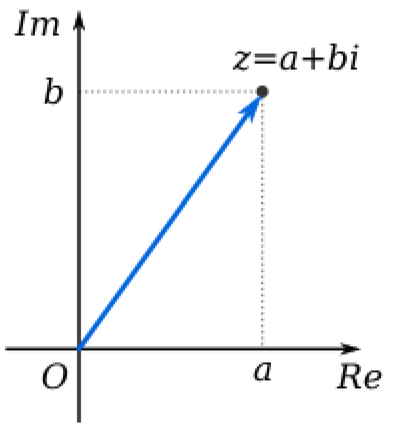
Object projected to hyperplane

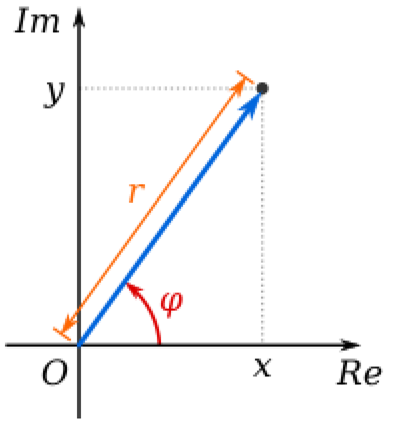
Will revisit when we discuss temporal embeddings and temporal KGs

- Many others: structured embedding (SE), FtransE, StransE, TransR, TransD, …
- We will focus on rotation and factorization, which work better

Relation as rotation

- Complex number $a + b\sqrt{-1} = a + jb$
- Another complex number $r = \cos\theta + j\sin\theta$
- Then $(a + jb)r$ rotates (a, b) in the complex plane by angle θ anticlockwise
- 💡 Model relation as rotation instead of translation
  - Elegant handling of anti/symmetry, inversion, composition
- Add more capacity (multiple complex dims) to handle many-to-many relations
- Close to top methods





RotatE

- Norm of complex number $a + jb$ is written as $|a + jb| = \sqrt{a^2 + b^2}$
- Let $\boldsymbol{s}, \boldsymbol{o} \in \mathbb{C}^D$ be complex vectors
- Norm of complex vector $\boldsymbol{c}$ is written as $||\boldsymbol{c}|| = \sqrt{\sum_d |c_d|^2} = \sqrt{\sum_d \Re(c_d)^2 + \Im(c_d)^2}$
- Let $\boldsymbol{r} \in \mathbb{C}^D$ with $|r_d| = 1 \ \forall d$
- $f(s, r, o) = ||\boldsymbol{s} \odot \boldsymbol{r} - \boldsymbol{o}||^2$

$$= \sum_d [\Re(s_d r_d - o_d)^2 + \Im(s_d r_d - o_d)^2]$$

Must ensure during gradient descent

Real part

Imaginary part

KG properties supported by RotatE

- RotatE can simulate TransE
- Relation r is symmetric if
  $$(s, r, o) \in \mathrm{KG} \Rightarrow (o, r, s) \in \mathrm{KG} \;\; \forall s, o$$
  - $\boldsymbol{o} = \boldsymbol{s} \odot \boldsymbol{r}$ and $\boldsymbol{s} = \boldsymbol{o} \odot \boldsymbol{r} \Rightarrow \;\boldsymbol{r} \odot \boldsymbol{r} = \mathbf{1}$
  - I.e., rotation $\boldsymbol{r}$ is its own inverse ⇒180° rotation
- Relation r is anti-symmetric if
  $$(s, r, o) \in \mathrm{KG} \Rightarrow (o, r, s) \notin \mathrm{KG} \; \forall s, o$$
  - For anti-symmetric relation choose a different angle

RotatE properties, continued

- Relations r and r' are inverses of each other if

$$(s, r, o) \in \mathrm{KG} \Rightarrow (o, r', s) \in \mathrm{KG} \;\; \forall s, o$$

  - Inversion modeled by complex conjugate: if r is represented as $\cos\theta + j\sin\theta$, then r^{-1} is represented as $\cos\theta - j\sin\theta$
- Composition of relations is equivalent to adding angles of rotation
  - $r_1 \mapsto \exp(j\theta_1), \;\; r_2 \mapsto \exp(j\theta_2) \Rightarrow$
    $r_1 \circ r_2 \mapsto \exp(j(\theta_1 + \theta_2))$
  - $(e_0, r_1, e_1), (e_1, r_2, e_2)$ means $\boldsymbol{e}_0 \odot \boldsymbol{r}_1 \odot \boldsymbol{r}_2 = \boldsymbol{e}_2$