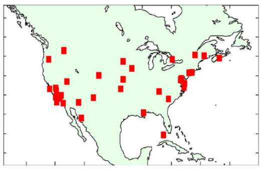
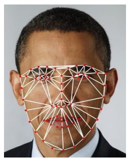
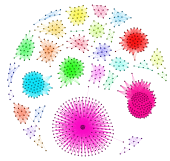
Optimal Regression on Sets

Zoltán Szabó[1], Bharath K. Sriperumbudur[2], Barnabás Póczos[3], Arthur Gretton[1] ([1]Gatsby Unit, UCL; [2]Department of Statistics, PSU; [3]Machine Learning Department, CMU).

- Example: aerosol prediction $\Rightarrow$ regression on labelled bags.

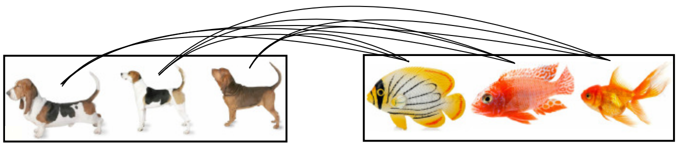
- Bag-of-objects representation: experimental design, computer vision, NLP, networks.

- Problem formulation:
  - Labelled training bags: $\hat{\mathbf{z}} = \left\{\left(\hat{P}_i, y_i\right)\right\}_{i=1}^{\ell}$. Test bag: $\hat{P}$. $N := |\hat{P}_i|$.
  - Estimator, prediction:

$$f_{\hat{\mathbf{z}}}^{\lambda} = \arg\min_{f \in \mathcal{H}(K)} \frac{1}{\ell} \sum_{i=1}^{\ell} \left[f\Big(\underbrace{\mu_{\hat{P}_i}}_{\text{feature of } \hat{P}_i}\Big) - y_i \right]^2 + \lambda \|f\|_{\mathcal{H}}^2,$$

$$\hat{y}\left(\hat{P}\right) = \mathbf{g}^T(\mathbf{G} + \ell\lambda\mathbf{I})^{-1}\mathbf{y},\ \mathbf{g} = \left[K\left(\mu_{\hat{P}}, \mu_{\hat{P}_i}\right)\right],\ \mathbf{G} = \left[K\left(\mu_{\hat{P}_i}, \mu_{\hat{P}_j}\right)\right],\ \mathbf{y} = [y_i].$$

  - Similarity of distributions: $K(P,Q) = \mathbb{E}_{a\sim P, b\sim Q} k(a,b) = \Big\langle \underbrace{\mathbb{E}_a \varphi(a)}_{\text{feature of distribution } P =: \mu_P}, \mathbb{E}_b \varphi(b) \Big\rangle$.



  - Quality of estimator: $\mathcal{R}(f) = \mathbb{E}_{(\mu_P, y)\sim\rho}[f(\mu_P) - y]^2$, f_ρ = best regressor.

- Known: assuming $f_\rho \in \mathcal{H}$
  - best/achieved rate: $\mathcal{R}(f_{\mathbf{z}}^{\lambda}) - \mathcal{R}(f_\rho) = \mathcal{O}\left(\ell^{-\frac{bc}{bc+1}}\right)$,
  - b – size of the input space, c – smoothness of f_ρ.

- **Our result**:
  - With $N = \tilde{\mathcal{O}}(\ell^a)$ and $2 \le a$, $f_{\hat{\mathbf{z}}}^{\lambda}$ has the best achievable rate; $a = \frac{b(c+1)}{bc+1} < 2$: ✓
  - $\Rightarrow$ Regression with set kernel is consistent (17-year-old open problem).
  - Practical: state-of-the-art aerosol prediction.

- Details (JMLR submission), code (in ITE):

```
http://arxiv.org/abs/1411.2066, https://bitbucket.org/szzoli/ite/
```

- Acknowledgements: This work was supported by the Gatsby Charitable Foundation, and by NSF grants IIS1247658 and IIS1250350. The work was carried out while Bharath K. Sriperumbudur was a research fellow in the Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics at the University of Cambridge, UK.