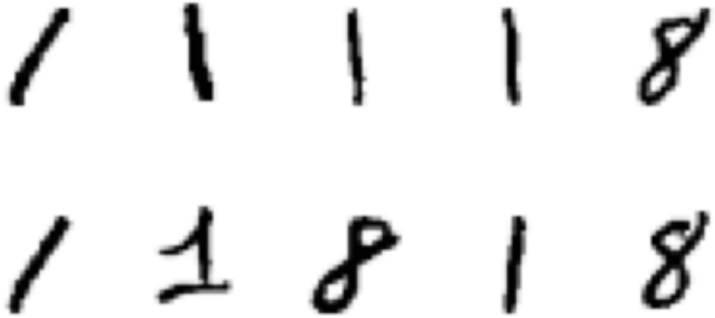
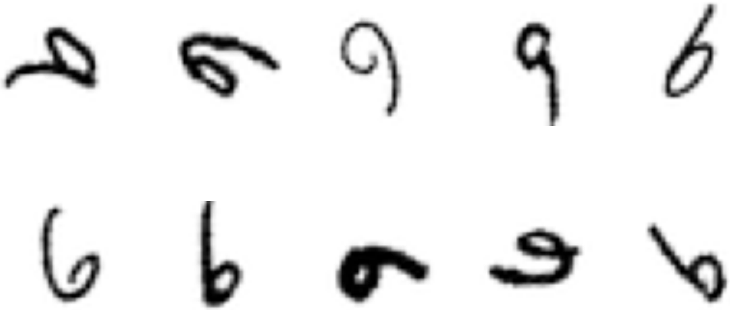
Studying the impact of assuming symmetries on learning

Or: few shot learning with symmetries

Andrea Perin and Stéphane Deny
Aalto University
SPAML, September 18, 2023





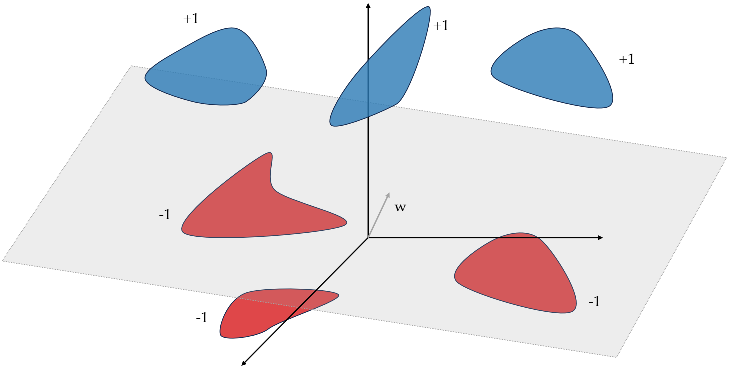
Symmetries in data

Natural data contain symmetries.
Effects of taking symmetry into account for classification:

- beneficial in some cases: **robustness**, out of distribution **generalisation**;
- harmful in some others: confusing digits, **loss of signal**.

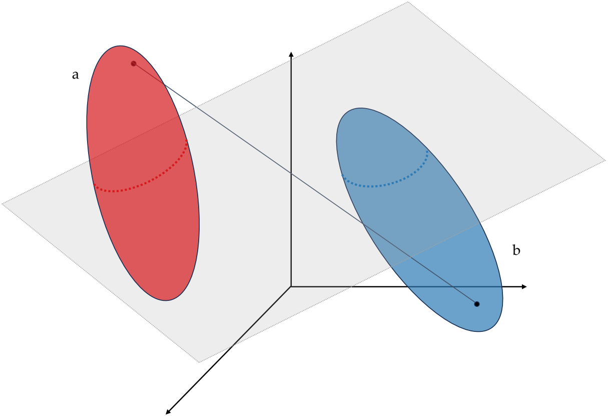
Question: Can we quantify the benefits/drawbacks of taking symmetries into account for classification?

Linear separability of manifolds



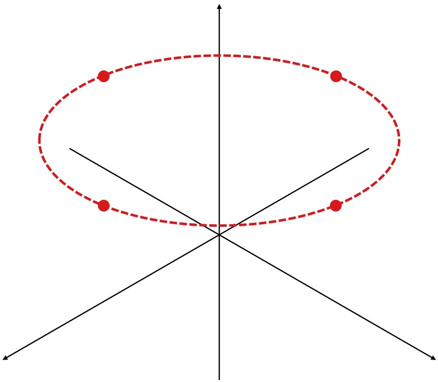
What are the conditions for linear separability of P manifolds?
Chung et al. (2018), Phys. Rev. X.

Few shot learning on manifolds



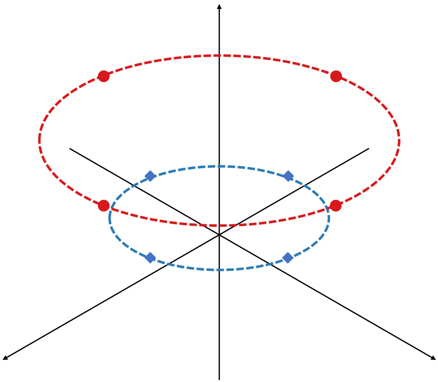
What is the error fraction of a few shot max margin linear separator?
Sorscher et al. (2021), bioRxiv.

Group structured linear separators



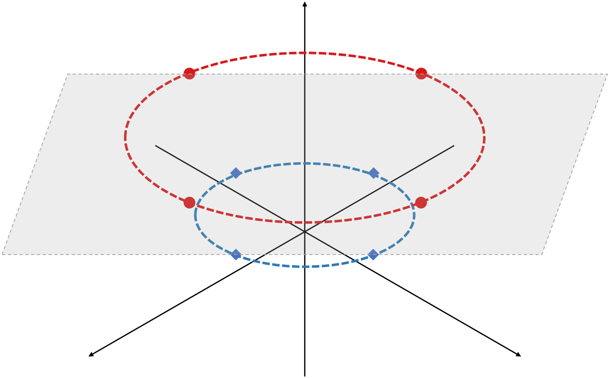
What is the capacity of group structured linear separators?
Farrell et al. (2022), arXiv.

Group structured linear separators



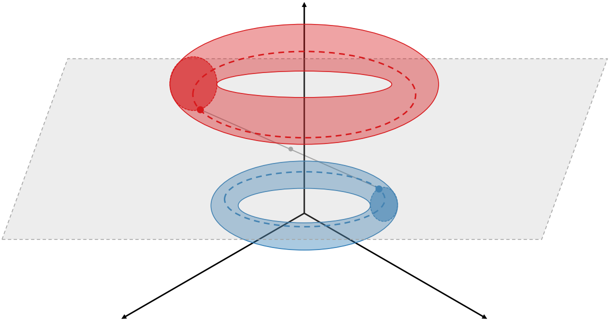
What is the capacity of group structured linear separators?
Farrell et al. (2022), arXiv.

Group structured linear separators



What is the capacity of group structured linear separators?
Farrell et al. (2022), arXiv.

Few shot learning on group structured data



Combining few shot learning with group structured manifolds.

The framework

Following *Sorscher et al. (2022)*:

- Binary classification of manifolds a and b:

$$x^a = x_0^a + \sum_i u_i^a R_i^a s_i^a, \quad x^b = x_0^b + \sum_i u_i^b R_i^b s_i^b \tag{1}$$

  with $s^a, s^b \sim \mathcal{U}(\mathbb{S}^{n-1})$.
- Maximum margin classification in *few shot* regime;
- In the presence of a *group action* operating on the data:

$$\rho(g) : V \to V, \quad x \mapsto \rho(g)x \tag{2}$$

  for a group G and a representation $\rho : G \to GL(\mathbb{R}, n)$;

What is the error fraction for the max margin separator?

Our results

In the framework we introduced, we derive the following results:

- the maximum margin separator is parallel to the orbits, and only uses invariant subspace information;
- the projection of the ellipsoidal manifolds on the invariant subspace is gaussian;
- we can rederive a formula for the error fraction in the gaussian case, as per Sorscher et al. (2021).

Group-induced split of data space

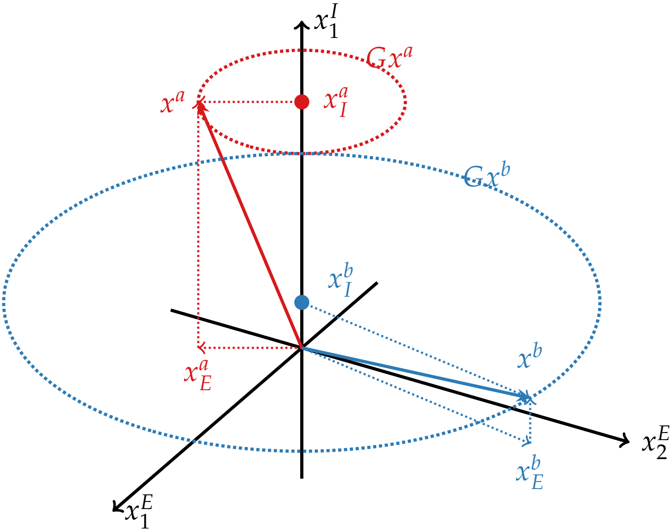
The group acts via linear representation on the data space.
Every point x is mapped to an orbit Gx:

$$Gx = \{\rho(g)x : g \in G\} \tag{3}$$

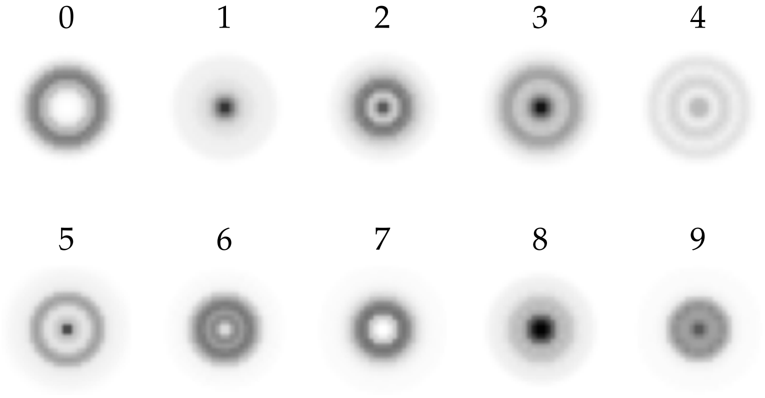
Group actions by linear representations induce a split into **invariant and equivariant subspaces**.
Invariant portion is left unchanged by group action.
We split the description of points:

$$x = x^I + x^E \tag{4}$$

Group-induced split of data space



Elements of the invariant subspace



Max margin separation and invariant subspace

When we average these orbits, we find that they collapse to **points** lying on the invariant subspace.
As a consequence, the maximum margin separator can only use information *on the invariant subspace;* we can restrict our analysis to this subspace.

How does the projection of a manifold on the invariant subspace look like?

Our results

In the framework we introduced, we derive the following results:

- the maximum margin separator is parallel to the orbits, and only uses invariant subspace information;
- the projection of the ellipsoidal manifolds on the invariant subspace is gaussian;
- we can rederive a formula for the error fraction in the gaussian case, as per Sorscher et al. (2021).

Projections become Gaussian

Starting n dimensional uniform ellipsoidal distribution:

$$f_n(x_1, x_2, \cdots, x_n) \propto \delta(x^T A x - 1). \tag{5}$$

After projecting k coordinates:

$$f_k(x) := f(x_{k+1}, \cdots, x_n) \propto \Theta(1 - \tilde{s}_k)(1 - \tilde{s}_k)^{\frac{k}{2} - 1}, \tag{6}$$

where $\tilde{s}^k$ is a quadratic form of the surviving x.
Approximation: when k is large, we can say

$$(1 - \tilde{s}_k)^{\frac{k-2}{2}} \approx \exp\left(-\left(\frac{k}{2} - 1\right)\tilde{s}_k\right), \tag{7}$$

and thus become approximately gaussian.

Our results

In the framework we introduced, we derive the following results:

- the maximum margin separator is parallel to the orbits, and only uses invariant subspace information;
- the projection of the ellipsoidal manifolds on the invariant subspace is gaussian;
- we can rederive a formula for the error fraction in the gaussian case, as per Sorscher et al. (2021).

Error fraction

The error fraction for manifold a is

$$\epsilon_a = \Pr_{x^a, x^b, \xi^a} \left[\left\| x^b - \xi^a \right\|^2 - \| x^a - \xi^a \|^2 < 0 \right], \tag{8}$$

where $x^a \in a$ and $x^b \in b$ are reference points, and $\xi^a \in a$ is the test point.
N.B.: asymmetric quantity!

In practice, it is computed by estimating the signal to noise ratio (SNR) of the manifolds, then computing the gaussian tail function of the SNR.

Error fraction

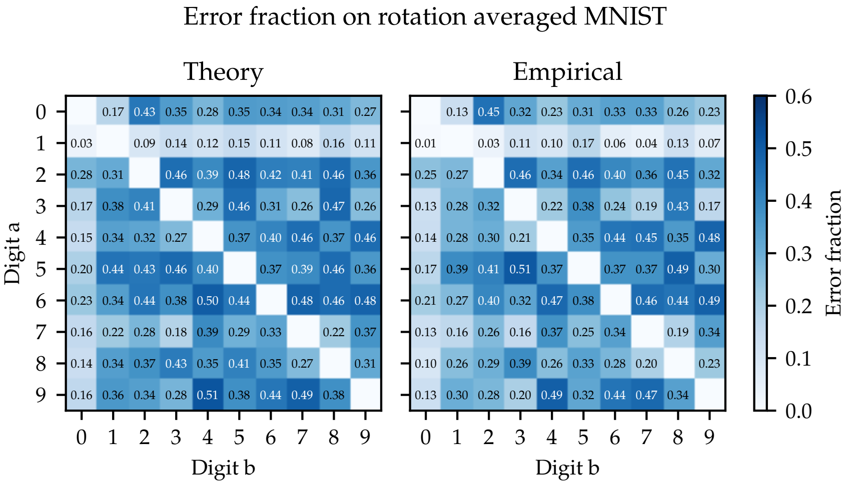
In the gaussian projected case, we find

$$\mathrm{SNR}_a = \frac{\|\Delta x_0\|^2 + \mathrm{tr}(\Sigma^b) - \mathrm{tr}(\Sigma^a)}{\sqrt{10\mathrm{tr}^2\Sigma^a + 2\mathrm{tr}^2\Sigma^b + 4\mathrm{tr}(\Sigma^a\Sigma^b) + \Delta x_0^T(\Sigma^a + \Sigma^b)\Delta x_0}}. \tag{9}$$

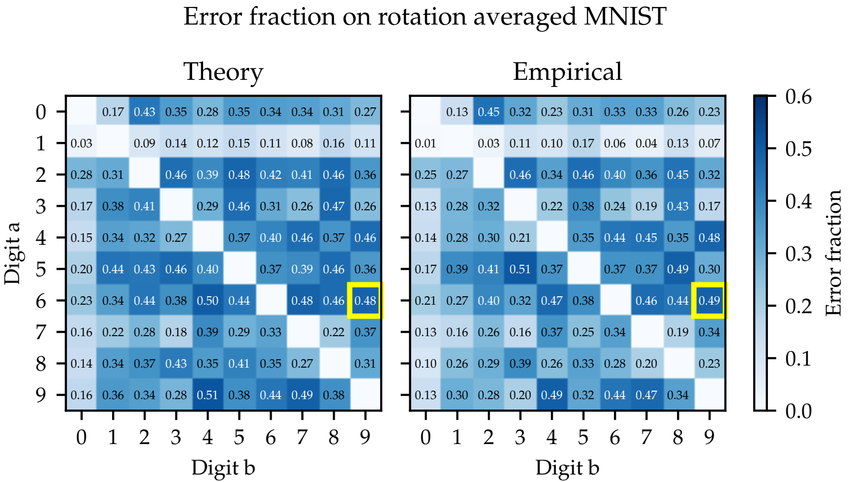
Sorscher et al. (2021)'s result:

$$\mathrm{SNR}_a = \frac{1}{2}\frac{\|\Delta x_0\|^2 + (R_b^2 R_a^{-2} - 1)}{\sqrt{D_a^{-1} + \|\Delta x_0 \cdot U_b\|^2 + \|\Delta x_0 \cdot U_a\|^2}}. \tag{10}$$

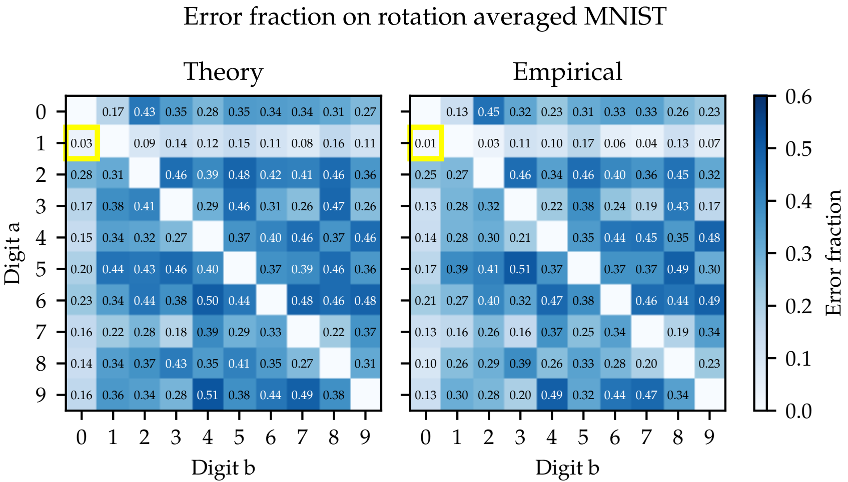
Error fraction - experiments



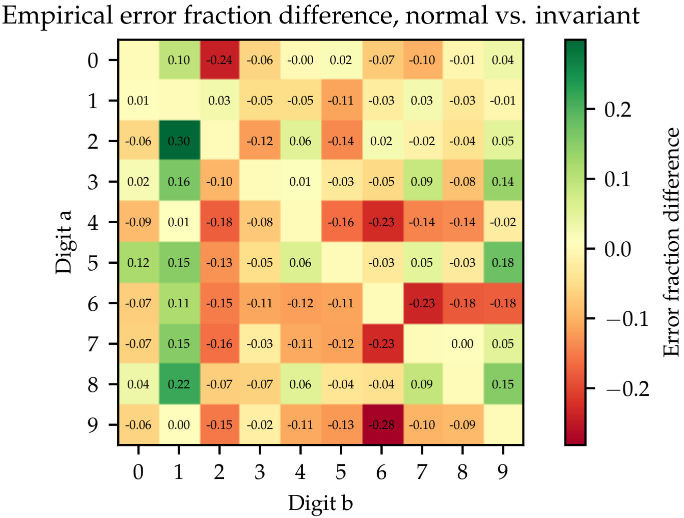
Error fraction - experiments



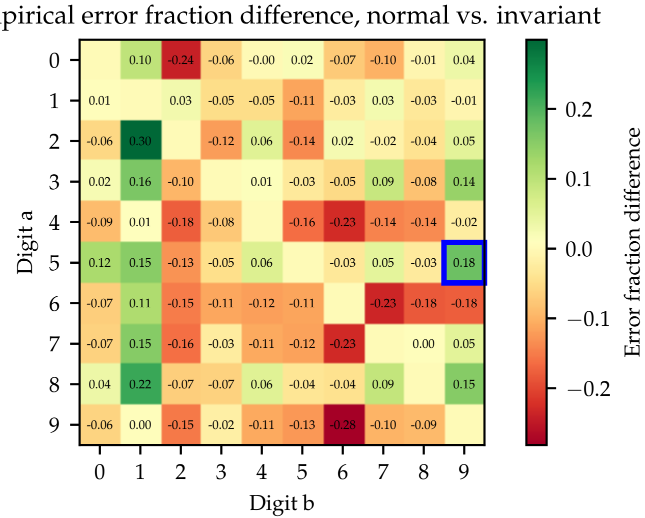
Error fraction - experiments



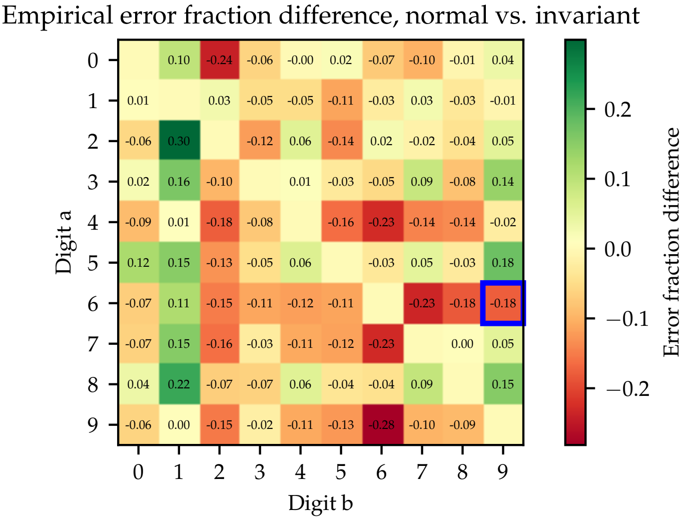
Error fraction - invariant vs. normal



Error fraction - invariant vs. normal



Error fraction - invariant vs. normal



Discussion

The inclusion of symmetries induces a **change in SNR**, and so on the error rate:

$$\epsilon_a \approx H(\mathrm{SNR}_a). \tag{11}$$

Intuitively:

- **beneficial** symmetries: decrease in variance **larger** than loss of signal;
- **harmful** symmetries: decrease in variance **smaller** than loss of signal.

Recap

How does taking symmetries into account impact classification?

- group action induces an invariant/equivariant split;
- the maximum margin separator uses only invariant information;
- we can project the manifolds on the invariant subspace;
- under this projection, ellipsoidal manifolds become gaussian;
- we can re-derive a formula for the error rate in the gaussian case.

Limitations and future steps

- linear separators are weak; how to extend to different learning algorithms?
- projection on the invariant subspace removes lots of signal: how to improve?
- transformations may not be uniformly distributed: how to generalise?

Recap

How does taking symmetries into account impact classification?

- group action induces an invariant/equivariant split;
- the maximum margin separator uses only invariant information;
- we can project the manifolds on the invariant subspace;
- under this projection, ellipsoidal manifolds become gaussian;
- we can re-derive a formula for the error rate in the gaussian case.