LINEAR LEARNERS

April 19, 2020

Contents

0 Overview **1**
0.1 Coordinates for Euclidean Learners 3

0 Overview

We begin our study of learners with one of the most ubiquitous learning algorithms: *linear regression.* in this context, we endow the label space $\mathfrak{y}$ with the structure of a finite-dimensional inner product space. $\mathfrak{y}$ is thus equipped with a norm in particular. Before we continue our study of linear regression we recall the following standard consructions of inner product spaces:

Lemma 0.1. *Let V and W be inner product spaces with orthonormal bases $\big(v_1, \dots v_m\big) \in V$ and $\big(w_1, \dots, w_n\big) \subset W$. Then*

1. *$V^* \stackrel{\text{def}}{=} \text{Hom}_{\mathbb{R}}(V, \mathbb{R})$ is an inner product space with $\langle f\,, g\,\rangle \stackrel{\text{def}}{=} \langle v\,, v'\,\rangle$ if $\langle v\,, -\,\rangle = f$ and $\langle -\,, v'\,\rangle = g$ and orthonormal basis given by the maps $\Big(\langle v_1\,, -\,\rangle, \dots, \langle v_m\,, -\,\rangle\Big)$*
2. *$V \times W$ is an inner product space with $\langle (v, w)\,, (v', w')\,\rangle \stackrel{\text{def}}{=} \langle v\,, w\,\rangle + \langle v'\,, w'\,\rangle$ and orthonormal basis $\Big((v_i, w_j)\Big)_{i,j}$*
3. *$V \otimes W$ is an inner product space with $\langle v \otimes w\,, v' \otimes w\,\rangle \stackrel{\text{def}}{=} \langle v\,, w\,\rangle \cdot \langle v'\,, w'\,\rangle$ and orthonormal basis $\Big((v_i \otimes w_j)\Big)_{i,j}$*
4. *$\text{Hom}(V, W)$ is an inner product space with an orthonormal basis given by maps $e_{i,j}$ defined as*

$$e_{i,j}(v_k) = \begin{cases} 0, & \text{if } k \neq i \\ w_j, & \text{otherwise} \end{cases} \tag{1}$$

Proof. Items (1) through (3) consist of routine calculations. To show item (4), recall that the function

$$\iota : V^* \otimes W \longrightarrow \text{Hom}_{\mathbb{R}}(V, W)$$

which assigns to $f \otimes w$ the linear map $\iota_{f \otimes w}(v) \stackrel{\text{def}}{=} f(v) \cdot w$ is an isomorphism. Now, items (1) and (3) together imply that $V^* \otimes W$ is indeed an inner product space with orthonormal basis $\Big\{\, \langle v_i\,, -\,\rangle \otimes w_j \Big\}_{i,j}$. Now simply note that the map associated to $\langle v_i\,, -\,\rangle \otimes w_j$ under the function ι is exactly $e_{i,j}$ X

Returning to our discussion above, we consider a finite-dimensional inner product space $\mathfrak{y}$ of labels and $\mathfrak{X}$ any set of features. Now, for any finite dataset $\Delta \subset \mathfrak{X} \times \mathfrak{y}$, the space $\mathfrak{y}^\Delta$ in turn carries an inner product by Lemma 0.1, so that for any map $f \in \mathfrak{y}^\mathfrak{X}$, we can define a cost as follows:

$$c(\Delta, f) = ||y - f(x)||_{\mathfrak{y}^\Delta} \stackrel{\text{def}}{=} \sqrt{\sum_\Delta ||y - f(x)||^2}$$

The main result of this chapter will be to show that this cost indeed describes a learner under the right conditions. To this end, we will make the following definition:

Definition 0.2. Let $\mathfrak{H} \subset \mathfrak{y}^\mathfrak{X}$ and $\Delta \subset \mathfrak{X} \times \mathfrak{y}$. Then we say that Δ separates $\mathfrak{H}$ if for any $f, g \in \mathfrak{H}$:

$$f|_{\Delta_\mathfrak{X}} = g|_{\Delta_\mathfrak{X}} \implies f = g$$

The main result of this chapter is:

Theorem 0.3. *Let $\mathfrak{y}$ be a finite-dimensional inner product space, $\mathfrak{X}$ any set and let $\mathfrak{H} \subset \mathfrak{y}^\mathfrak{X}$ be a finite-dimensional subspace of $\mathfrak{y}^\mathfrak{X}$. Assume that any $\Delta \in \mathfrak{D}$ separates $\mathfrak{H}$ and let $c(\Delta, f) = ||y - f(x)||_{\mathfrak{y}^\Delta}$. Then $(\mathfrak{X}, \mathfrak{y}, \mathfrak{D}, \mathfrak{H}, c)$ defines a sharp learner (as in Definition **??**)*

After having reviewed the necessary linear algebraic, we now have the ingredients to prove the promised main theorem of logistic regression, which we recall below for the reader's convenience:

Theorem 0.4 (linear regression)**.** *Let $\mathfrak{y}$ be a finite dimensional inner product space, $\mathfrak{X}$ any set. Let $\mathfrak{H} \subset \mathfrak{y}^\mathfrak{X}$ be a finite dimensional subspace and*

$$\mathfrak{D} = \left\{ \Delta \subset \mathfrak{X} \times \mathfrak{y} \;\middle|\; |\Delta| \neq \infty \text{ and } \Delta \text{ separates } \mathfrak{H} \right\}.$$

and

$$c : \mathfrak{D} \times \mathfrak{H} \longrightarrow \mathfrak{R} : (\Delta, f) \mapsto ||(y - f(x))_{(x,y)\in\Delta}||_{\mathfrak{y}^\Delta}$$

Then $\left(\mathfrak{X}, \mathfrak{y}, \mathfrak{D}, \mathfrak{H}, c\right)$ is a sharp learner.

Proof. Let $\Delta \in \mathfrak{D}$.
Since $\mathfrak{H}$ is finite dimensional, we can choose an inner product on it. Moreover, the space $\mathfrak{y}^\Delta$ is canonically a finite dimensional inner product space by Lemma 0.1
We now consider the map:

$$\text{ev}_\Delta : \mathfrak{H} \longrightarrow \mathfrak{y}^\Delta : f \mapsto f(\Delta_\mathfrak{X})$$

(where we used notation **??**). It is easy to see that this map is linear. Moreover, the fact that Δ separates $\mathfrak{H}$ is equivalent to ev_Δ being injective
WE note now that ev_Δ allows us to rewrite the cost function $c(\Delta, f)$ as follows:

$$c(\Delta, f) = ||\big(y - f(x)\big)_{(x,y)\in\Delta}||_{\mathfrak{y}^\Delta} = ||\Delta_\mathfrak{y} - \text{ev}_\Delta(f)||_{\mathfrak{y}^\Delta}$$

It follows that f that minimizes the cost $c(\Delta, f)$ iff f minimizes the distance between $\Delta_\mathfrak{y}$ and $\text{ev}_\Delta(f)$. By Lemma **??**, this is in turn equivalent to requiring that $\text{ev}_\Delta(f)$ is the projection of the vector $\Delta_\mathfrak{y}$ onto the image of the map $\text{ev}_\Delta : \mathfrak{H} \longrightarrow \mathfrak{y}^\Delta$. We can now describe this image using Moore-Penrose inverses: indeed, let

$$h_\Delta \stackrel{\text{def}}{=} \text{ev}_\Delta^+(\Delta_\mathfrak{y})$$

Then Lemma **??** implies that $\text{ev}_\Delta(f)$ is the projection of $\Delta_\mathfrak{y}$ onto the image of ev_Δ^+ if and only if f lies in the affine subspace $h_\Delta + \ker(\text{ev}_\Delta) \subset \mathfrak{y}^\Delta$. We now recall that ev_Δ is in fact injective so that finally $f = h_\Delta$ X

As a corollary, we can show that the Euclidean learner introduced in Definition 0.12 indeed is a learner provide an explicit formula for the learned hypothesis h_Δ

Corollary 0.5. *Let $\mathfrak{X}$ and $\mathfrak{y}$ be finite dimensional inner product spaces,*

$$\mathfrak{D} = \left\{ \Delta \subset \mathfrak{X} \times \mathfrak{y} \;\middle|\; |\Delta| \neq \infty, \text{ and } span(\Delta_{\mathfrak{X}}) = \mathfrak{X} \right\}, \; \mathfrak{H} = \mathrm{Hom}_{\mathbb{R}}(\mathfrak{X}, \mathfrak{y})$$

and

$$c : \mathfrak{D} \times \mathfrak{H} \longrightarrow \mathbb{R} : (\Delta, f) \mapsto ||\big(y - f(x)\big)_{(x,y)\in\Delta}||_{\mathfrak{y}^\Delta}$$

Then $(\mathfrak{X}, \mathfrak{y}, \mathfrak{D}, \mathfrak{H}, c)$ is a sharp linear learner. Moreover,

$$h_\Delta = \mathrm{ev}_\Delta^+ \left(\Delta_{\mathfrak{y}}\right) = \left(\left(\mathrm{ev}_\Delta^* \circ \mathrm{ev}_\Delta \right)^{-1} \circ \mathrm{ev}_\Delta^* \right)(\Delta_{\mathfrak{y}})$$

where $\mathrm{ev}_\Delta : \mathfrak{H} \longrightarrow \mathfrak{y}^\Delta : f \mapsto f(\Delta_{\mathfrak{y}})$

Proof. By Theorem 0.3, we simply need to note that $\mathfrak{H} = \mathrm{Hom}_{\mathbb{R}}(\mathfrak{X}, \mathfrak{y}$ is finite dimensional and that each dataset Δ indeed separates $\mathfrak{H}$ since $\Delta_{\mathfrak{X}}$ spans the whole of $\mathfrak{X}$, so that f is fully characterized on $\Delta_{\mathfrak{X}}$. The formula for h_Δ follows from the proof of Theorem 0.3, where we showed that $h_\Delta = ev^+(\Delta_y)$ and corollary **??** since ev_Δ is injective X

0.1. Coordinates for Euclidean Learners

In the last section of this chapter, we will once and for all fix a Euclidean learner $\mathfrak{L}$ (as defined in 0.12) and introduce coordinates on the feature space $\mathfrak{X}$ and label space $\mathfrak{y}$. This will allow us to represent the learned hypothesis h_Δ by a matrix. We will show that this matrix indeed coincides with what is classically referred to as the *regression matrix*.
Let us pick orthonormal bases $\mathcal{E} \stackrel{\text{def}}{=} (v_1, \ldots, v_m)$ for $\mathfrak{X}$ and $\mathcal{F} \stackrel{\text{def}}{=} (w_1, \ldots w_n)$ for $\mathfrak{y}$. We will denote the associated coordinate maps by $\mathrm{co}_{\mathcal{E}}$ and $\mathrm{co}_{\mathcal{F}}$ respectively.
Now to any map $f \in \mathrm{Hom}_{\mathbb{R}}(\mathfrak{X}, \mathfrak{y})$, we'll associate the matrix $M_f \in \mathrm{Mat}_{n\times m}(\mathbb{R})$ whose i-th column is given by $\mathrm{co}_{\mathcal{F}}(v_i)$. It is well known that this matrix satisfies

$$\mathrm{co}_{\mathcal{F}}\big(f(v)\big) = M_f \cdot \mathrm{co}_{\mathcal{E}}(v)$$

Our goal in this section is to compute the matrix M_{h_Δ} associated to the learned hypothesis h_Δ of the dataset Δ. More precisely, we will prove the following theorem:

Theorem 0.6. *Let $\Delta = \big((x_1, y_1), \ldots, (x_d, y_d)\big) \in \mathfrak{D}$ be a dataset.*
Let $X = [\mathrm{co}_{\mathcal{E}}(x_i)_i] \in \mathrm{Mat}_{m\times d}(\mathbb{R})$ and $Y = [\mathrm{co}_{\mathcal{F}}(y_j)_j] \in \mathrm{Mat}_{d\times n}(\mathbb{R})$.
Then the matrix corresponding to the learned hypothesis $h_\Delta \in \mathrm{Hom}_{\mathbb{R}}(\mathfrak{X}, \mathfrak{y})$ is given by

$$M_{h_\Delta} = Y \cdot X^+,$$

*where X^+ is the Moore-Penrose pseudo-inverse of X (see **??**).*
In other words, for any feature $v \in \mathfrak{X}$, we have

$$\mathrm{co}_{\mathcal{F}}\big(h_\Delta(v)\big) = \big(Y \cdot X^+\big) \cdot \mathrm{co}_{\mathcal{E}}(v)$$

We will prove this using a series of lemma's.
Since $h_\Delta = \mathrm{ev}_\Delta(\Delta_{\mathfrak{y}})$ by corollary 0.5, it's worth reinterpreting the map $\mathrm{ev}_\Delta : \mathrm{Hom}(\mathfrak{X}, \mathfrak{y}) \longrightarrow \mathfrak{y}^d$ as a map between matrix spaces through the use of the orthonormal bases $\mathcal{E}$ and $\mathcal{F}$. To this end, we consider the isomorphism $M : \mathrm{Hom}(\mathfrak{X}, \mathfrak{y}) \longrightarrow \mathrm{Mat}_{n\times m}(\mathbb{R})$which assigns to f its matrix representation M_f, after the choice of bases $\mathcal{E} \subset \mathfrak{X}$ and $\mathcal{F} \subset \mathfrak{y}$. We will also consider the map α_P the multiplication by a matrix P on the right

Lemma 0.7. *Let X denote the matrix $\left[\operatorname{co}_{\mathcal{E}}(x_i)_i\right]$. Then the diagram:*

$$\begin{array}{ccc} \operatorname{Hom}(\mathfrak{X},\mathfrak{y}) & \xrightarrow{\operatorname{ev}_\Delta} & \mathfrak{y}^d \\ {\scriptstyle M_{(-)}}\downarrow & & \downarrow{\scriptstyle (\operatorname{co}_{\mathcal{F}})^d} \\ \operatorname{Mat}_{n\times m}(\mathbb{R}) & \xrightarrow[\alpha_X]{} & \operatorname{Mat}_{d\times n}(\mathbb{R}) \end{array}$$

is commutative

Proof. This is really just a restatement of the various definitions:
Let $f \in \operatorname{Hom}(\mathfrak{X},\mathfrak{y})$. Then $\left(ev_\Delta \circ \operatorname{co}_{\mathcal{F}}^d\right)(f)$ is the matrix $\left[(\operatorname{co}_{\mathcal{F}}(f(x_i))_i\right]$, whose i-th column is the vector $\operatorname{co}_{\mathcal{F}}\left(f(x_i)\right) \in \mathbb{R}^n$.
Going around the other way, we obtain the matrix $\alpha_X \circ M_f = M_f \cdot X$, whose i-th column is $M_F \cdot \left[\operatorname{co}_{\mathcal{E}}(x_i)\right]$.
Now, by the definition of M_f, we have $M_f \cdot [\operatorname{co}_{\mathcal{E}}(x_i)] = \operatorname{co}_{\mathcal{F}}\left(f(x_i)\right)$, proving the claim

X

Next, we will use the above lemma to describe the Moore-Penrose pseudo-inverse of ev_Δ in terms of the maps M, α_X and $\left(\operatorname{co}_{\mathcal{F}}\right)^d$. We'll need a little result from Euclidean geometry to do this.. recall that a map f on an inner product space V is *orthogonal* if it preserves the inner product:

$$\langle u\,,v\,\rangle = \langle f(u)\,,f(v)\,\rangle$$

This is equivalent to the adjoint coinciding with the inverse: $f^* = f^{-1}$. We now have the following linear algebraic lemma:

Lemma 0.8. *Let V and W be finite dimensional inner product spaces.*
Let $f \in \operatorname{Hom}(V,W)$ and let $\phi \in \operatorname{Hom}_{\mathbb{R}}(V,V)$ and $\psi \in \operatorname{Hom}_{\mathbb{R}}(W,W)$ be orthogonal maps. Then the Moore-Penrose inverse of $\psi \circ f \circ \phi \in \operatorname{Hom}(V,W)$ is given by

$$(\psi \circ f \circ \phi)^+ = \phi^{-1} \circ f^+ \circ \psi^{-1}$$

Proof. We need to check the 4 conditions of criterium (3) in Lemma **??**. First

$$(\phi^{-1} \circ f^+ \circ \psi^{-1}) \circ (\psi \circ f \circ \phi) \circ (\phi^{-1} \circ f^+ \circ \psi^{-1}) = (\phi^{-1} \circ f^+ \circ f \circ f^+ \circ \psi^{-1}) = \phi^{-1} \circ f^+ \circ \psi^{-1}$$

The second condition is analogous.
To prove the third condition, we invoke that ϕ and ψ are orthogonal, so that $\phi^{-1^*} = \phi$ and $\psi^{-1^*} = \psi$. We now compute:

$$\left((\phi^{-1} \circ f^+ \circ \psi^{-1}) \circ (\psi \circ f \circ \phi)\right)^* = \left(\phi^{-1} \circ f^+ \circ f \circ \phi\right)^* = \phi^* \circ (f^+ \circ f)^* \circ \phi^{-1^*} = \phi \circ (f^+ \circ f) \circ \phi$$

Where the last line follows from the orthogonality of ϕ and the fact that $f^+ \circ f$ is self-adjoint since f^+ is the Moore-Penrose pseudo-inverse to f.
The fourth condition is proven analogously.

X

To use the above lemma we'll need to explain how $\operatorname{co}_{\mathcal{F}}$ and M are indeed orthogonal maps. For the benefit of the reader we first recap the inner products involved in the commutative diagram of Lemma 0.7:

- The bilinear form on $\operatorname{Hom}(\mathfrak{X},\mathfrak{y})$ does not have a general description, however Lemma 0.1 implies that the maps

$$e_{i,j}(v_k) = \begin{cases} 0, & \text{if } k \neq i \\ w_j, & \text{otherwise} \end{cases} \tag{2}$$

  form an orthonormal basis

- The space $\mathfrak{y}^d$ has the inner product defined by $\langle (y_1, \dots, y_n), (y'_1, \dots y'_n) \rangle = \sum_i \langle y_i, y'_i \rangle$ and orthonormal basis $\left\{(w_{i_1}, \dots w_{i_m})\right\}_{i_1, \dots i_m}$ by Lemma 0.1
- The matrix spaces are endowed with the *Frobenius inner product*: for matrices A and B, it's defined as $\langle A, B \rangle = \sum_{i,j} A_{i,j} \cdot B_{i,j}$. It follows immediately that the elementary matrices $\left\{E_{i,j}\right\}_{i,j}$ form an orthonormal basis for these spaces.
  Another description of this inner product will be useful later on. Indeed, we have:

$$\mathrm{Tr}(A \cdot B^t) = \sum_k (A \cdot B^t)_{k,k} = \sum_k (\sum_j A_{k,j} B^t_{j,k}) = \sum_k (\sum_j A_{k,j} B_{k,j}) = \langle A, B \rangle$$

  This allows us to prove certain properties. For example, for any matrix P, we have

$$\langle A, B \cdot P \rangle = \mathrm{Tr}\left(A \cdot (B \cdot P)^t\right) = \mathrm{Tr}(A \cdot P^t \cdot B^t) = \left\langle (A \cdot P^t), B \right\rangle$$

We now have:

Lemma 0.9. *The maps M and $(\mathrm{co}_{\mathcal{F}})^d$ are orthogonal*

Proof. Indeed, it suffices to show that both maps M and $\left(\mathrm{co}_{\mathcal{F}}\right)^d$ send an orthonormal basis to an orthonormal basis. Now, the map M sends the orthonormal basis $\left\{e_{i,j}\right\}_{i,j}$ to the orthonormal basis of elementary matrices $\left\{E_{i,j}\right\}_{i,j}$.
The map $(\mathrm{co}_{\mathcal{F}})^d$ in turn sends the orthonormal basis $\left\{(w_{i_1}, \dots w_{i_m})\right\}_{i_1, \dots i_m}$ to the elementary matrices $\left\{E_{i,j}\right\}_{i,j}$ as well. X

Combining both lemma's hence lets us write the Moore-Penrose inverse of ev^+_Δ as

$$\mathrm{ev}^+_\Delta = \left(M^{-1} \circ \alpha_X \circ \mathrm{co}^d_{\mathcal{F}}\right)^+ = M^{-1} \circ \left(\alpha_X\right)^+ \circ \mathrm{co}^d_{\mathcal{F}}$$

We now wish to simplify the map $\left(\alpha_X\right)^+$. To this end we note that for any matrix $P \in \mathrm{Mat}_{d \times m}(\mathbb{R})$, we can associate its Moore-Penrose inverse as the inverse following example **??**

Lemma 0.10. *For any $P \in \mathrm{Mat}_{d \times m}(\mathbb{R})$, let α_P denote the right multplication by P*

$$\alpha_P : \mathrm{Mat}_{n \times m}(\mathbb{R}) \longrightarrow \mathrm{Mat}_{d \times n}(\mathbb{R}) : A \mapsto A \cdot P$$

Then we have $\left(\alpha_P\right)^+ = \alpha_{P^+}$

Proof. We simply need to show that α_{P^+} satisfies the conditions of Moore-Penrose pseudo-inverse by checking (3) of Lemma **??**.
The first two conditions are absolutely trivial.
To prove the third condition, we need to show that the map $\alpha_P \circ \alpha_{P^+}$ (which coincides with $\alpha_{P^+ \cdot P}$) is self-adjoint. To this end, we let $A, B \in \mathrm{Mat}_{n \times m}(\mathbb{R})$. The discussion before Lemma 0.9 yields:

$$\langle A, \alpha_{P^+ \cdot P}(B) \rangle = \mathrm{Tr}\left(A \cdot \left(B \cdot P^+ \cdot P\right)^t\right) = \mathrm{Tr}\left(A \cdot (P^+ \cdot P)^t \cdot B\right) = \mathrm{Tr}\left(A \cdot (P^+ \cdot P) \cdot B^t\right) = \langle \alpha_{P^+ \cdot P}(A), B \rangle$$

Where the 3rd equality follows from the fact that the matrix $P^+ \cdot P$ is symmetric by Lemma **??**.
The proof of condition (4) is completely analogous. X

The above lemma allows us to rewrite the Moore-Penrose pseudo-inverse one step further

$$\mathrm{ev}^+_\Delta = M^{-1} \circ \alpha_{X^+} \circ \left(\mathrm{co}_{\mathcal{F}}\right)^d$$

We can now prove the main theorem of this section:

proof of Theorem 0.6. We want to show that $M_{h_\Delta} = Y \cdot X^+$. Now, we know from Corollary 0.5 that $h_\Delta = \mathrm{ev}^+_\Delta\left(\Delta_{\mathfrak{y}}\right) = \mathrm{ev}^+_\Delta\left(y_1, \dots y_d\right)$. Moreover, the above discussion shows that

$$\mathrm{ev}^+_\Delta = M^{-1} \circ \alpha_{X^+} \circ \left(\mathrm{co}_{\mathcal{F}}\right)^d$$

So that applying the set of features $(y_1, \dots y_d)$ to the left hand side followed by the map M yields

$$M_{h_\Delta} = \left(\alpha_{X^+} \circ \mathrm{co}^d_{\mathcal{F}}\right)(y_1, \dots, y_d) = \alpha_{X^+}\left(\left[\mathrm{co}_{\mathcal{F}}(y_1) \dots \mathrm{co}_{\mathcal{F}}(y_d)\right]\right) = \alpha_{X^+}(Y) = Y \cdot X^+$$

as required X

Definition 0.11. We say that a learner is linear if it is of the above form

One particular type of linear learner will be of particular interest, as it allows us to be a little more explicit with certain constructions: if we assume that $\mathfrak{X}$ itself carries the structure of a finite-dimensional inner product space, and put $\mathfrak{H} \stackrel{\mathrm{def}}{=} \mathrm{Hom}_{\mathbb{R}}(\mathfrak{X}, \mathfrak{y})$ as well as consider finite datasets $\Delta \subset \mathfrak{X} \times \mathfrak{y}$ such that $\Delta_{\mathfrak{X}}$ spans the whole of $\mathfrak{X}$ (where we recall our use of the notation **??**), then it's easy to see that the conditions of Theorem 0.3 are satisfied so that we indeed obtain an example of a linear learner:

Definition 0.12. A *Euclidean learner* is a sharp (linear) learner where $\mathfrak{X}, \mathfrak{y}$ are finite-dimensional inner product spaces,

$$\mathfrak{D} = \left\{\Delta \subset \mathfrak{X} \times \mathfrak{y} \;\middle|\; |\Delta| \neq \infty, \text{ and } \mathrm{span}(\Delta_{\mathfrak{X}}) = \mathfrak{X}\right\},\ \mathfrak{H} = \mathrm{Hom}_{\mathbb{R}}(\mathfrak{X}, \mathfrak{y})$$

and

$$c : \mathfrak{D} \times \mathfrak{H} \longrightarrow \mathbb{R} : (\Delta, f) \mapsto ||\big(y - f(x)\big)_{(x,y)\in\Delta}||_{\mathfrak{y}^\Delta}$$