J. R. Landers

Forecast-Weighted Drift Regularization

Use case and motivation. A frozen predictor may face covariates with different temporal dynamics: one sensor ages steadily, another varies seasonally, and a third remains stable. Historical trajectories can support a separate velocity forecast for each coordinate. The proposed extension uses these forecasts to concentrate sensitivity penalties where future motion is expected, including its magnitude and uncertainty. The intended targets are nuisance changes, such as calibration drift; suppressing sensitivity to genuinely predictive changes can harm accuracy.

Theoretical context. The Jacobian–velocity framework [1] studies absolutely continuous covariate paths $X_t \in \mathbb{R}^d$, $t \in [0,T]$, and a fixed scalar predictor f_θ. Let g_θ be a fixed conditional performance field and $r(t) = \mathbb{E}[g_\theta(X_t)]$. Under its chain-rule and integrability assumptions, and directional domination $|\nabla g_\theta(X_t)^\top \dot{X}_t| \le \beta |\nabla f_\theta(X_t)^\top \dot{X}_t|$, the temporal variance satisfies

$$\mathcal{V}_T := \frac{1}{T}\int_0^T (r(t)-\bar{r})^2 dt \le \frac{\beta^2 T}{\pi^2}\int_0^T \mathbb{E}[(\nabla f_\theta(X_t)^\top \dot{X}_t)^2]dt, \qquad \bar{r} := \frac{1}{T}\int_0^T r(t)dt. \tag{1}$$

Existing DTR penalizes sensitivity in a selected drift subspace. Tangent Prop [2] and learned manifold tangents [3] establish the broader precedent for directional derivative penalties; selective input-gradient penalties [4] supply a coordinate-wise analogue. Here the proposed addition is a forecast-dependent velocity metric and an explicit account of its estimation error, rather than a new chain-rule mechanism.

Mathematical extension. Condition on predeployment training data, so all fitted functions below are fixed. Fit coordinate forecasts $\hat{b}_j(t,x)$ for velocity and assemble $\hat{b} \in \mathbb{R}^d$. Let $\hat{\Sigma}(t,x) \succeq 0$ estimate conditional velocity covariance, and define

$$\widehat{M}(t,x) = \hat{b}(t,x)\hat{b}(t,x)^\top + \hat{\Sigma}(t,x), \qquad \mathcal{R}_{\widehat{M}}(\theta) = \frac{1}{T}\int_0^T \mathbb{E}[\nabla f_\theta(X_t)^\top \widehat{M}(t,X_t)\nabla f_\theta(X_t)]dt. \tag{2}$$

Train with supervised loss plus $\lambda \mathcal{R}_{\widehat{M}}$, approximated using historical or forecast trajectories. The true conditional second moment $M(t,x) := \mathbb{E}[\dot{X}_t \dot{X}_t^\top \mid X_t = x]$ recovers the integrand of (1) exactly through the tower property. If, almost surely for almost every deployment time, the fitted metric satisfies $M(t,X_t) \preceq \kappa \widehat{M}(t,X_t) + \varepsilon^2 I$ for constants $\kappa \ge 1$, $\varepsilon \ge 0$, then

$$\boxed{\mathcal{V}_T \le \frac{\beta^2 T^2}{\pi^2}\left(\kappa \mathcal{R}_{\widehat{M}}(\theta) + \varepsilon^2 \mathcal{R}_I(\theta)\right), \qquad \mathcal{R}_I := \frac{1}{T}\int_0^T \mathbb{E}\|\nabla f_\theta(X_t)\|^2 dt.} \tag{3}$$

This follows by applying the matrix inequality to the gradient in (1). The calibration condition is an additional hypothesis requiring evidence; fitting drift forecasts alone does not establish it. The bound concerns deployment expectations, so replacing them by training averages also requires control of that approximation.

Interpretation and scope. Using only $\mathrm{diag}(\widehat{M})$ yields separate penalties $\sum_j \widehat{M}_{jj}(\partial_j f_\theta)^2$; the full matrix retains coordinated motion and cancellation. A diagonal approximation needs its own calibration guarantee. Equation (3) separates forecast-aligned control from an isotropic safeguard for missed motion. Uncertainty matters because mean velocities alone can cancel while squared velocity remains large. Estimating velocities requires trajectories or an explicit temporal coupling; marginal distributions alone do not identify them. If directional domination instead includes a conditional-risk remainder q_t, the bound becomes twice the right side of (3) plus $2T\pi^{-2}\int_0^T \mathbb{E}q_t^2 dt$. Temporal stability does not by itself ensure low average risk.

References. [1] J. R. Landers, *Jacobian-Velocity Bounds for Deployment Risk Under Covariate Drift*, 2026; current manuscript. [2] P. Simard et al., "Tangent Prop," *NeurIPS 4*, 1992. [3] S. Rifai et al., "The Manifold Tangent Classifier," *NeurIPS 24*, 2011.

[4] A. S. Ross, M. C. Hughes, and F. Doshi-Velez, "Right for the Right Reasons," *IJCAI*, pp. 2662–2670, 2017.